

Explanation-based learning in infancy

Renée Baillargeon¹ · Gerald F. DeJong²

Published online: 11 July 2017
© Psychonomic Society, Inc. 2017

Abstract In explanation-based learning (EBL), domain knowledge is leveraged in order to learn general rules from few examples. An explanation is constructed for initial exemplars and is then generalized into a candidate rule that uses only the relevant features specified in the explanation; if the rule proves accurate for a few additional exemplars, it is adopted. EBL is thus highly efficient because it combines both analytic and empirical evidence. EBL has been proposed as one of the mechanisms that help infants acquire and revise their physical rules. To evaluate this proposal, 11- and 12-month-olds ($n = 260$) were taught to replace their current support rule (that an object is stable when half or more of its bottom surface is supported) with a more sophisticated rule (that an object is stable when half or more of the entire object is supported). Infants saw teaching events in which asymmetrical objects were placed on a base, followed by static test displays involving a novel asymmetrical object and a novel base. When the teaching events were designed to facilitate EBL, infants learned the new rule with as few as two (12-month-olds) or three (11-month-olds) exemplars. When the teaching events were designed to impede EBL, however, infants failed to learn the rule. Together, these results demonstrate that even infants, with their limited knowledge about the world, benefit from the knowledge-based approach of EBL.

Keywords Infant cognition · Knowledge acquisition · Explanation-based learning

Introduction

Infants acquire a large number of rules that identify relevant features for predicting the outcomes of occlusion, containment, collision, support, and other physical events (for reviews, see Baillargeon, Li, Gertner, & Wu, 2011; Baillargeon et al., 2012). These rules are general and are strikingly similar across infants. Yet, for any given rule, (a) each infant observes a unique and relatively small set of events from which to extract the rule, and (b) each event includes numerous potential features. How, then, do infants acquire these rules?

We have proposed that *explanation-based learning* (EBL; DeJong, 1993, 2014) is one of the processes that enable infants to efficiently acquire and revise their physical rules (e.g., Baillargeon et al., 2011; Wang & Baillargeon, 2008). The EBL process has three main steps. The first is *triggering*: When infants encounter outcomes they cannot explain on the basis of their current rules, the EBL process is triggered. In situations in which no existing rule applies, infants may notice unexplained variation in the events' outcomes; in situations in which an existing rule does apply, infants may notice that although some outcomes support the rule, others contradict it. Either way, exposure to the unexplained outcomes triggers EBL.

The second step in the EBL process is *explanation construction and generalization*: Infants bring to bear their physical-domain knowledge (i.e., their core knowledge and previously acquired rules; e.g., Baillargeon, 2008; Baillargeon & Carey, 2012; Baillargeon, Li, Ng, & Yuan, 2009; Baillargeon, Wu, Yuan, & Luo, 2009; Carey, 2009; R. Gelman, 1990; Keil, 1995; Leslie, 1995; Spelke, 1994; Spelke, Breinlinger, Macomber, & Jacobson, 1992) to construct a plausible

✉ Renée Baillargeon
rbailiar@illinois.edu

¹ University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, IL 61820, USA

² Department of Computer Science, University of Illinois at Urbana-Champaign, 201 N Goodwin Avenue, Urbana, IL 61801, USA

explanation for the outcomes they have observed. Though rarely correct from a physicist's perspective, this explanation still provides a rudimentary causal analysis that specifies which features of the events contributed to their outcomes—other features are implicitly omitted. As such, the explanation is easily generalized, resulting in a candidate rule that incorporates only the relevant features specified in the explanation.

The final step in the EBL process is *empirical confirmation*: Once a rule has been hypothesized, it must be evaluated against further empirical evidence, which will serve to either confirm or reject it. If the candidate rule proves accurate in predicting outcomes for a few additional exemplars, it is adopted and becomes part of infants' domain knowledge. From then on, it guides their predictions and actions (e.g., Hespos & Baillargeon, 2006; Wang & Kohne, 2007) and can also be recruited in explanations for other events.

Two points about the EBL process deserve emphasis. First, this process makes clear (a) why infants generally acquire similar rules, even though each infant experiences a unique and relatively small set of events with many potential features, and also (b) why infants generally do not acquire rules based on specious or accidental regularities in their environments. In each case, infants' domain knowledge constrains the rules that are acquired, because only regularities that can be plausibly explained are adopted as rules. In the field of statistical machine learning, by contrast, distinguishing specious from genuine patterns constitutes a major problem, known as *overfitting* (e.g., Bishop, 2006; Mitchell, 1997; Murphy, 2012). Mathematically, the number of possible patterns grows combinatorially with the number of (observable and derivable) features available. Thus, with a limited number of examples and a myriad of features in each example, many specious patterns can fit the data. Ruling out these patterns statistically requires an exponentially large number of examples.

This brings us to the second point. The EBL process also makes clear why infants may require only a few exemplars to acquire a new rule. Because EBL combines *both* analytic evidence (i.e., the explanation that is constructed for the observed events and then generalized into a candidate rule) *and* empirical evidence, it makes possible highly efficient learning.

Prior teaching experiments with infants

Our EBL account not only describes how infants acquire their physical rules: It also suggests how infants might be “taught” a rule they have not yet acquired, via exposure to EBL-designed observations. To evaluate this suggestion, previous experiments (Wang & Baillargeon, 2008; Wang & Kohne, 2007) attempted to teach 9-month-olds a rule about covering events that is typically not acquired until about 12 months (Wang & Baillargeon, 2006; Wang, Baillargeon, & Paterson, 2005): When a rigid cover (or upside-down container) is placed over

an object, their relative heights determine whether the object will become fully or only partly hidden.

Infants received three pairs of teaching trials. In each pair, a tall and a short cover (differing only in height) were lowered over a tall object; infants could observe that the object became fully hidden under the tall cover, but remained partly visible beneath the short cover. Different covers were used in the three teaching pairs. Following these trials, infants detected a violation when a tall object became fully hidden under a short cover (Wang & Baillargeon, 2008), and they correctly searched for a tall object under a tall as opposed to a short cover (Wang & Kohne, 2007), suggesting that they had acquired the rule.

From an EBL perspective, these results are readily interpretable. First, during the teaching trials, infants noticed unexplained variation in the events' outcomes (the object became sometimes fully hidden and sometimes only partly hidden), which triggered the EBL process. Second, infants brought to bear their physical-domain knowledge to generate an explanation for these differential outcomes: The principle of persistence (Baillargeon, 2008; Baillargeon, Li, et al., 2009) dictated that because the object continued to exist and retained its height when under a cover, it could become fully hidden only under the tall cover. Third, infants received sufficient empirical evidence to confirm the rule suggested by their explanation: All three pairs of teaching covers behaved in accordance with the rule. Infants therefore adopted the rule, enabling them to succeed at violation-of-expectation and manual-search tasks involving new covers and objects.

Further results supported this analysis. Consistent with the EBL account, infants failed to acquire the rule if the teaching object was much shorter and became fully hidden under the short as well as the tall cover in each teaching pair; there was then no unexplained variation in outcomes that could trigger EBL. Infants also failed to acquire the rule if the teaching covers were shown to have false bottoms that rendered them all very shallow; although the tall teaching object still became fully hidden under the tall cover and partly hidden under the short cover, infants could no longer generate a plausible explanation for these outcomes.

Development of infants' knowledge about support events

To provide converging evidence that EBL underlies infants' rapid acquisition and revision of their physical rules, in the present research we conducted teaching experiments focused on support events involving inert objects (henceforth simply *objects*).¹ In

¹ Support events involving self-propelled objects or animate objects have somewhat different rules. For example, when a novel self-propelled object is released in midair, young infants do not detect a violation if the object remains suspended, presumably because they endow the object with internal energy and infer that the object is using its energy to resist falling (e.g., Baillargeon, Wu, et al., 2009; Leslie, 1995; Luo, Kaufman, & Baillargeon, 2009; Setoh, Wu, Baillargeon, & Gelman, 2013). In this article, we focus on simple, everyday support events in which an inert object is released on an inert base.

these experiments, we attempted to lead infants to replace an existing support rule with a more sophisticated one. Before introducing our experiments, we briefly discuss some of the core principles that contribute to early reasoning about support, as well as some of the support rules that infants acquire in the first year of life.

Core principles At least two core principles guide early reasoning about support events. One principle is *gravity*: Objects fall when unsupported (Baillargeon, Wu, et al., 2009; Luo, Kaufman, & Baillargeon, 2009; Needham & Baillargeon, 1993; Wang et al., 2005). The other principle is *persistence*: All other things being equal, objects persist, with their properties, in time and space (Baillargeon, 2008; Baillargeon, Li, et al., 2009; Spelke et al., 1992; Spelke, Phillips, & Woodward, 1995). The principle of persistence has multiple corollaries, but the one most relevant to support events is *solidity*: A solid object cannot pass through space occupied by another solid object (e.g., Baillargeon & DeVos, 1991; Baillargeon, Spelke, & Wasserman, 1985; Hespos & Baillargeon, 2001; Spelke et al., 1992).

Support rules At 2.5–4 months of age, infants expect an object to fall when it is released in midair (e.g., Baillargeon, 1995; Baillargeon, Wu, et al., 2009; Luo et al., 2009; Needham & Baillargeon, 1993). When an object is released in contact with a base, however, infants have no particular expectation as to whether the object will remain stable or fall: Their representation of the event (“object released in contact with base”) is too sparse or imprecise for their core knowledge to generate a prediction about the object’s stability. Nevertheless, infants observe some unexplained variation in outcomes as objects released in contact with bases sometimes remain stable and sometimes fall. These unexplained observations trigger EBL, leading to the acquisition, at about 4.5–5 months of age, of a *location-of-contact* rule: An object is stable when it is released on top of, but not against, a base (e.g., Baillargeon, 1995; Hespos & Baillargeon, 2008; Needham & Baillargeon, 1997). The principles of gravity and solidity provide a ready explanation for this rule: When an object is released on top of a base, the base effectively blocks the object’s fall, because the object cannot pass through the base; when an object is released against a base, however, there is nothing to block the object’s fall. This first rule thus serves to establish a new event category, “support” (or more specifically, “passive support from below”), which describes a causal interaction between two objects with distinct event roles: A “support” blocks the fall of a “supportee.”

Over time, infants come to recognize that their location-of-contact rule is imperfect; although some outcomes are consistent with the rule, others are not, because objects sometimes fall even when released on top of bases. Exposure to these unexplained outcomes again triggers EBL. By about

6.5 months of age, infants replace their location-of-contact rule with a new *proportion-of-contact* rule: When released on top of a base, an object remains stable as long as half or more of its bottom surface rests on the base (e.g., Baillargeon, Needham, & DeVos, 1992; Dan, Omori, & Tomiyasu, 2000; Hespos & Baillargeon, 2008; Huettel & Needham, 2000; Luo et al., 2009; Wang, Zhang, & Baillargeon, 2016). Infants are thus learning to attend not only to whether an object has been released on top of a base, but also to how much of the object actually rests on the base. Infants’ initial focus on the contact between the object’s bottom surface and the base could be due to a number of factors: First, this contact is where the base blocks the object’s fall; second, because many of the objects that young infants encounter in everyday life are symmetrical, attending to what proportion of the object’s bottom surface lies on the base initially provides an easy proxy for predicting the object’s stability. When this proportion is less than half, infants consider the object to be inadequately supported and expect it to fall.²

In the months that follow, infants come to realize that their proportion-of-contact rule is in need of revision; here again, although some outcomes are consistent with the rule, others are not. In particular, as infants’ motor skills improve, they become more likely to encounter asymmetrical objects, and hence to observe outcomes that contradict their proportion-of-contact rule: Objects sometimes fall when released with half or more of their bottom surfaces supported. Exposure to these unexplained outcomes triggers EBL and leads to the acquisition, by about 13 months of age, of a new *proportional-distribution* rule: When released with one end on a base, an object remains stable as long as half or more of the entire object rests on the base (e.g., Baillargeon, 1995, 1998, 1999). Thus, when an asymmetrical object is released with one end on a base, infants attend to what proportion of the object as a whole (not just of its bottom surface) lies on the base. If this proportion is less than half, infants consider the object to be inadequately supported and expect it to fall.³

² Support for this analysis comes from errors of commission that infants with a proportion-of-contact rule produce (errors of commission occur when infants detect violations in events that are physically possible but happen to contradict infants’ imperfect rules; Luo & Baillargeon, 2005). For example, 7.5-month-olds detect a violation when a rectangular box remains stable with only the middle third of its bottom surface supported on a narrow base; because less than half of the box’s bottom surface rests on the base, infants expect the box to fall, and they (mistakenly) detect a violation when it remains stable instead (Dan et al., 2000; Wang et al., 2016).

³ The proportional-distribution rule is, of course, still partly incorrect and can lead to false predictions. For example, infants would expect an L-shaped box with equally large vertical and horizontal portions to remain stable with the horizontal portion off the base, and they would (mistakenly) detect a violation if the box fell, thus producing an error of commission. Attention to distance information appears to be a late accomplishment: In solving balance-scale problems, for example, 5-year-olds typically consider the weights on each side of the scale, but not the distance of the weights from the fulcrum (Siegler 1976; Siegler & Chen, 1998).

The present research

In the present research, we attempted to teach the proportional-distribution rule to 12-month-olds (Exp. 1) and 11-month-olds (Exps. 2 and 3). All infants saw the same two static test displays, which involved a yellow L-shaped box and a blue rectangular base (Fig. 1). In each display, the right half of the box's bottom surface lay on the base; what varied was the box's orientation. In the *unexpected* display, the box looked like a typical L, and its smaller end was supported; in the *expected* display, the box looked like a backward L, and its larger end was supported. Prior to seeing these displays, all infants had received teaching trials. Some infants received appropriate teaching trials that were designed to foster the three critical steps in the EBL process, whereas other infants received inappropriate teaching trials that were designed to disrupt one or more of these steps. We reasoned that if the infants who had received appropriate teaching trials succeeded in learning the proportional-distribution rule, they would look reliably longer at the unexpected display, because it violated this rule. Moreover, if the infants who had received inappropriate teaching trials failed to learn the proportional-distribution rule, they would look equally at the two displays, because both were consistent with infants' proportion-of-contact rule. Together, these findings would provide strong support for our account of the EBL process.

Experiment 1

Twelve-month-olds were assigned to one of four conditions ($n = 20$ per condition in all experiments). Prior to seeing the two static test displays, infants received two pairs of teaching trials that varied across conditions.

Conditions

Two-exemplar condition Each pair of teaching trials in the two-exemplar condition involved a large-on event and a small-on event (Fig. 2a). At the start of the *large-on* event, an experimenter's (E) right gloved hand reached through a curtain in the left wall of a puppet-stage apparatus and held the smaller end of an asymmetrical box about 5 cm above and to the left of a red rectangular base. To start, the hand placed the right half of the box' bottom surface (i.e., the box's larger end) on the base (2 s), tapped the box on the base four times (2 s), paused with the box on the base (1 s), released the box and withdrew to its starting position (2 s), and paused (1 s). Next, the hand grasped the box (1 s), returned it to its starting position (2 s), and paused (1 s), ready to start a new 12-s event cycle. Cycles were repeated until the trial ended (see the Procedure section for criteria). The *small-on* event was identical, except that the box's orientation was reversed;

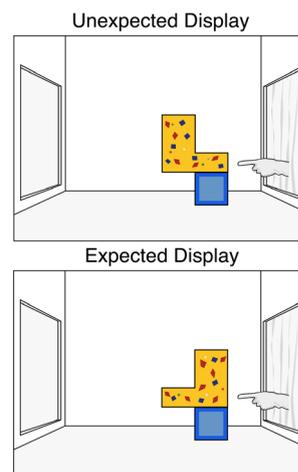


Fig. 1 Schematic depiction of the static test displays in Experiments 1–3.

consequently, E now placed the box's smaller end on the base, and the box fell when released. Different boxes were used in the two teaching pairs. Half of the infants saw a pink box shaped like a B on its back (henceforth, the B-box) in the first teaching pair and a green right-triangle box (henceforth, the T-box) in the second teaching pair; the other infants saw a pink T-box in the first teaching pair and a green T-box in the second teaching pair (this within-condition manipulation did not affect the test responses).

The teaching trials were designed to facilitate the three steps in the EBL process. First, in each teaching pair, the small-on event contradicted the infants' proportion-of-contact rule: The box fell even though half of its bottom surface rested on the base. Analysis of the teaching trials suggested that infants detected this violation: On average, they looked reliably longer at the small-on than at the large-on events (Table 1). The unexplained outcomes of the small-on events were expected to trigger EBL.

Second, because in each teaching pair the small-on and large-on events differed only in the box's orientation, infants were likely to focus on this difference in their quest for an explanation of the events' contrastive outcomes. By bringing to bear their physical-domain knowledge (as discussed earlier), infants could arrive at a plausible explanation for why the box fell in one orientation but not the other. Specifically, when the box was oriented in such a way that the proportion of the box on the base was smaller than that off the base, the box was then inadequately supported; the base could not passively block the box's fall when less than half of the entire box lay on the base. This explanation would lead infants to hypothesize the proportional-distribution rule: An object released with one end resting on a base will remain stable as long as the proportion of the entire object on the base is greater than that off the base.

Finally, infants received empirical evidence for this hypothesized rule because, across the teaching trials, they saw two different boxes behave in accordance with the rule. From

Experiment 1: Teaching Events

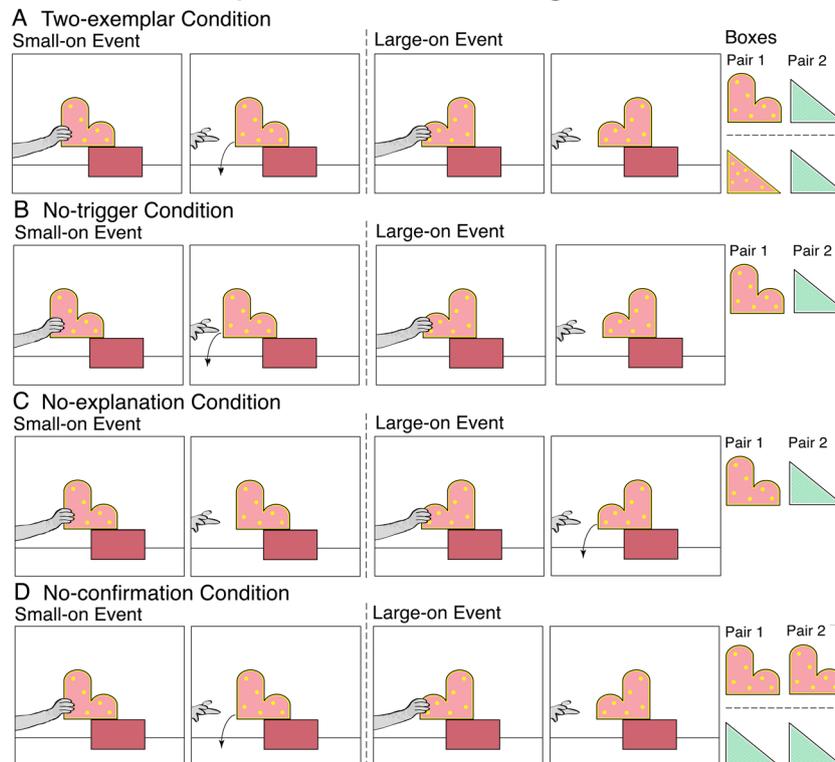


Fig. 2 Schematic depiction of the teaching events in each condition of Experiment 1. Infants received two teaching pairs; the events in the first teaching pair are depicted. In most conditions, the second teaching pair involved a different box; the boxes used in each teaching pair are depicted at the end of each row. In the two-exemplar and no-confirmation conditions, half of the infants saw the two boxes above the dashed line, and half

saw the boxes below the dashed line (right column). In the no-trigger condition, half of the infants saw large-on events in which the box was released with only 25% of its bottom surface supported (as shown); for the other infants, E first placed the box with 50% of its bottom surface on the base, but then lifted and tilted the box toward herself before releasing it (not shown).

a purely statistical perspective, two exemplars would seem to provide woefully insufficient confirmatory evidence for a new rule. Two exemplars can be sufficient in EBL, however, because the bulk of the evidence is analytic and derives from the plausibility of the explanation.

If infants were led by the teaching trials to replace their proportion-of-contact rule with the more sophisticated proportional-distribution rule, then they should apply this new rule to the test displays and look reliably longer at the unexpected than at the expected display. Note that these displays were designed to look superficially different from the teaching events: They were presented on the opposite side of the apparatus, they involved a novel box and base, and they were static (E simply pointed to the box with her gloved left hand, from a distance of about 4 cm). Nevertheless, an abstract proportional-distribution rule should enable infants to detect the violation in the unexpected display.

No-trigger condition The teaching trials in the no-trigger condition were identical to those in the two-exemplar condition, with two exceptions. First, in the small-on events, the box now fell for reasons consistent with infants' current knowledge

about support (Fig. 2b). For half of the infants, E placed only the right 25% of the box's bottom surface on the base; for the other infants, E placed the right 50% of the box's bottom surface on the base, as in the two-exemplar condition, but she lifted the box off the base and tilted it toward herself before releasing it (this within-condition manipulation did not affect the test responses). Second, all infants saw the pink B-box in the first teaching pair and the green T-box in the second teaching pair.

In this condition, infants never observed unexplained outcomes that could trigger EBL; their proportion-of-contact rule explained why the box fell in each small-on event and why it remained stable in each large-on event. Indeed, analysis of the teaching trials indicated that infants looked about equally at the small-on and large-on events, suggesting that they viewed them all as expected.⁴ Thus, even though the box still fell in each

⁴ The finding that infants in the no-trigger condition looked about equally at the small-on and large-on events is important. It suggests that infants in the two-exemplar and no-confirmation conditions looked reliably longer at the small-on events not simply because the falling boxes drew their attention, but because these events violated their proportion-of-contact rule. In other words, infants produced an error of commission, by viewing as unexpected events that were physically possible but that happened to contradict their imperfect rule.

Table 1 Mean looking times (and standard deviations) at the teaching events, separately per experiment and condition

	Small-On Event	Large-On Event	<i>F</i> value	<i>P</i> value	Cohen's <i>d</i>
Experiment 1: 12-month-olds					
Two-exemplar condition	48.7 (10.2)	40.8 (11.3)	10.88	.004	0.74
No-trigger condition	46.8 (11.3)	43.2 (14.0)	0.97	.337	0.28
No-explanation condition	39.9 (10.8)	49.0 (12.5)	7.44	.013	−0.78
No-confirmation condition	46.2 (12.6)	38.1 (11.2)	6.84	.017	0.68
Further Results: 11-month-olds					
Two-exemplar condition	50.0 (10.6)	38.5 (14.7)	29.49	.000	0.90
Experiment 2: 11-month-olds					
Three-exemplar condition	41.1 (10.1)	35.2 (10.0)	5.33	.032	0.59
No-trigger condition		36.3 (8.4)			
No-explanation condition	37.2 (10.9)	44.5 (10.8)	7.49	.013	−0.67
No-confirmation condition	50.9 (7.6)	42.0 (7.8)	16.67	.001	1.15
Experiment 3: 11-month-olds					
Three-exemplar condition	38.9 (9.8)	32.6 (9.0)	4.63	.044	0.67
Different-bases condition	47.3 (8.6)	39.2 (13.0)	12.50	.002	0.73
Different-boxes condition	42.9 (11.7)	27.0 (6.7)	44.09	.000	1.67
No-comparison condition					
Only small-on events	42.5 (9.4)				
Small-on then large-on events	43.4 (7.8)	21.8 (6.7)	53.75	.000	2.97

All conditions had 20 infants. In the no-trigger condition of Experiment 2, infants saw large-on events in two blocks of trials. In the no-comparison condition of Experiment 3, ten infants saw small-on events in two blocks of trials, and ten infants saw small-on events in a first block and large-on events in a second block.

small-on event, the EBL account predicted that infants would fail to revise their proportion-of-contact rule and hence would look equally at the unexpected and expected test displays.

No-explanation condition The teaching trials in the no-explanation condition were identical to those in the two-exemplar condition, with two exceptions. First, the teaching events had *reverse* outcomes (Fig. 2c): The box remained stable in the small-on events and fell in the large-on events. Second, all infants saw the pink B-box in the first teaching pair and the green T-box in the second teaching pair.

In each large-on event, infants observed an unexplained outcome that could trigger EBL: The box fell even though half of its bottom surface was supported, thus violating infants' proportion-of-contact rule. Analysis of the teaching trials suggested that infants detected this violation: On average, they looked reliably longer at the large-on than at the small-on events. By comparing these events, infants could determine that the main difference between them was the box's orientation. However, because the box now remained stable when released with its smaller end on the base and fell when released with its larger end on the base, infants could no longer use their physical-domain knowledge to generate a plausible explanation, thereby derailing the EBL process. Infants should thus fail to revise their proportion-of-contact rule, and hence should look equally at the two test displays.

No-confirmation condition The teaching trials in the no-confirmation condition were identical to those in the two-exemplar condition, except that infants saw the same asymmetrical box in both teaching pairs (Fig. 2d); half of the infants saw the pink B-box, and the other half saw the green T-box (this within-condition manipulation did not affect the test responses).

In each teaching pair, infants saw an unexplained outcome that could trigger EBL: In the small-on event, the box fell even though half of its bottom surface rested on the base, thereby contradicting infants' proportion-of-contact rule. Analysis of the teaching trials suggested that infants detected this violation: On average, they looked reliably longer at the small-on than at the large-on events. As in the two-exemplar condition, infants could recruit their physical-domain knowledge to construct a plausible explanation for the outcomes of the small-on events and hypothesize a proportional-distribution rule. However, because the same box was used in both teaching pairs, there was insufficient empirical evidence to confirm the rule. Moreover, the very next box infants encountered, the L-box in the test displays, provided *disconfirming* evidence for the rule: The box remained stable in the unexpected display, even though its larger end was off the base. The EBL account thus predicted that infants would discard the hypothesized rule and look equally at the unexpected and expected displays.

Method

Participants

The participants were 80 healthy term 12-month-olds (40 male, 40 female; $M = 11$ months, 20 days (11;20), range = 11;11–11;28), 20 per condition. Another 14 infants, distributed across conditions, were excluded because they were overly fussy or active (seven), looked the maximum allowed at both test displays (two), or had a mean looking-time difference between the two teaching events over three standard deviations from the condition mean (one). The remaining four infants were excluded for being inattentive during the teaching trials. Because we were attempting to teach infants a new rule, those with little interest in the teaching events were eliminated using the following criterion. Across conditions, infants watched, on average, 7.0–7.5 event cycles per teaching pair (out of a maximum of ten); thus, in this and the remaining experiments, infants were judged to be inattentive if they watched fewer than 3.75 event cycles per teaching pair.

Apparatus and stimuli

The apparatus consisted of a brightly lit display booth (126 cm high \times 102 cm wide \times 56 cm deep) mounted 76 cm above the room floor, with a large opening (51 \times 95 cm) in its front wall; between trials, a supervisor lowered a white curtain in front of this opening. Inside the apparatus, the back wall and floor were covered with pastel adhesive paper; an added layer under the floor helped reduce the noises caused by the boxes' fall. Each side wall was painted white and had a window (51 \times 38 cm) located 7 cm from the back wall, which was filled with either a fringe white curtain (when used by E) or a solid white curtain (when not).

The base in the teaching events was a red rectangular box (15 \times 27 \times 8 cm); it was centered in front of and positioned 37.5 cm from the left window. The asymmetrical boxes in the teaching events included a light pink B-box (26 \times 27 \times 8 cm) that was decorated with large yellow dots and outlined with yellow tape; a T-box (24 \times 27 \times 8 cm) identical in pattern and color to the B-box; and a light green T-box of identical size that was decorated with a small white dot pattern and outlined with white tape. Weighted copies of the pink B-box and green T-box were used in the no-explanation condition.

The base in the test displays was a light blue rectangular box (15 \times 13.5 \times 13.5 cm) outlined with blue tape; it was centered in front of and 24 cm from the right window. The box in the expected display was a yellow L-box (24 \times 27 \times 8 cm; smaller end alone: 8 \times 13.5 \times 8 cm) that was decorated with stars and red and blue quadrilaterals and was outlined with yellow tape; it was positioned 3.5 cm from the front edge of the base. An identical weighted box was used in the unexpected display.

E wore a long black glove on her right hand and a long silver glove on her left hand; she sat at the left window in the teaching trials and the right window in the test trials. During the session, a metronome beat softly to help E adhere to the events' second-by-second scripts. An image of the events was projected onto a television set located behind the apparatus and monitored by the supervisor to confirm that the events followed the prescribed scripts.

Procedure

Infants sat on a parent's lap centered in front of the left edge of the base used in the teaching events (to facilitate comparison of the portions of the box on and off the base); the parents were instructed to remain silent and to close their eyes during the test trials. Before the session, infants were shown E's gloved hands as well as the (nonweighted) boxes and bases to be used in the session, one at a time. Half of the infants saw the small-on event first in each teaching pair and the unexpected display first in the test trials; the other infants saw the large-on event first in each teaching pair and the expected display first in the test trials. During the test trials, the infant's looking behavior was monitored by two hidden observers; looking times were computed using the primary observer's responses. During the teaching trials, the primary observer was absent from the room and thus was naïve about both the condition to which the infant was assigned and the order in which the test displays were presented. Interobserver agreement in each test trial was calculated by dividing the number of 100-ms intervals during which the two observers agreed by the number of intervals in the trial. Agreement for all infants in this report averaged 92% per trial per infant.

Each teaching trial ended when infants had (a) looked away for two consecutive seconds after having looked for at least 12 cumulative seconds or (b) looked for 60 cumulative seconds. The 12-s minimal value ensured that infants had the opportunity to see at least one event cycle before a trial could end. Each test trial ended when infants had (a) looked away for one consecutive second after having looked for at least two cumulative seconds or (b) looked for 40 cumulative seconds. Because the test trials were static, infants tended to look away sooner, so smaller values were used.

Preliminary analyses of the test data in this report revealed no significant interaction of condition and event with either sex or order; the data were therefore collapsed across the latter two factors.

Results and discussion

Infants' test looking times (Fig. 3) were compared by an analysis of variance (ANOVA) with condition (two-exemplar, no-trigger, no-explanation, or no-confirmation) as a between-

subjects factor and display (unexpected or expected) as a within-subjects factor. The analysis yielded only a significant Condition \times Display interaction, $F(3, 76) = 3.14$, $p = .030$, $\eta_p^2 = .11$. Planned comparisons revealed that the infants in the two-exemplar condition looked reliably longer at the unexpected ($M = 16.6$, $SD = 10.1$) than at the expected ($M = 9.1$, $SD = 7.5$) display, $F(1, 76) = 8.14$, $p = .006$, Cohen's $d = 0.85$, whereas the infants in the no-trigger [unexpected: $M = 12.0$, $SD = 9.6$; expected: $M = 15.4$, $SD = 9.7$; $F(1, 76) = 1.68$, $p = .199$, $d = -0.36$], no-explanation [unexpected: $M = 10.3$, $SD = 6.4$; expected: $M = 11.0$, $SD = 9.4$; $F(1, 76) < 1$, $d = -0.09$], and no-confirmation [unexpected: $M = 13.0$, $SD = 9.8$; expected: $M = 11.1$, $SD = 10.1$; $F(1, 76) < 1$, $d = 0.19$] conditions looked about equally at the two displays.

The results of Experiment 1 supported the EBL account. In the two-exemplar condition, the teaching trials triggered and facilitated the EBL process, leading infants to replace their proportion-of-contact rule with a more sophisticated proportional-distribution rule. During the test trials, infants applied this new rule and thus detected the violation in the unexpected display, about one month before infants typically do so. In the other three conditions, the teaching events disrupted one or more of the critical steps in the EBL process. As a result, infants did not revise their proportion-of-contact rule and detected no violation in the unexpected display, which was consistent with this rule.

Further results with 11-month-olds

Encouraged by the positive results of the 12-month-olds in the two-exemplar condition, we next tested 20 11-month-olds in the same condition (11 male, nine female; $M = 11;4$, range = 10;18–11;10); another three infants were excluded because they were inattentive in the teaching trials (two) or had a mean looking-time difference between the two teaching events over three standard deviations from the condition mean (one). During the teaching trials, infants looked reliably longer at the small-on than at the large-on events, suggesting that they realized that the small-on events contradicted their proportion-of-contact rule. Nevertheless, infants failed to revise this rule and looked about equally at the unexpected ($M = 11.1$, $SD = 5.7$) and expected ($M = 10.4$, $SD = 7.2$) test displays, $F(1, 19) < 1$, $d = 0.11$. Their responses differed reliably from those of the 12-month-olds in the two-exemplar condition, $F(1, 38) = 4.36$, $p = .044$, $\eta_p^2 = .10$.

Experiment 2

Because two exemplars were insufficient to teach 11-month-olds the proportional-distribution rule, in Experiment 2 we increased this number to three exemplars. It seemed plausible that younger infants might require (a) more information to

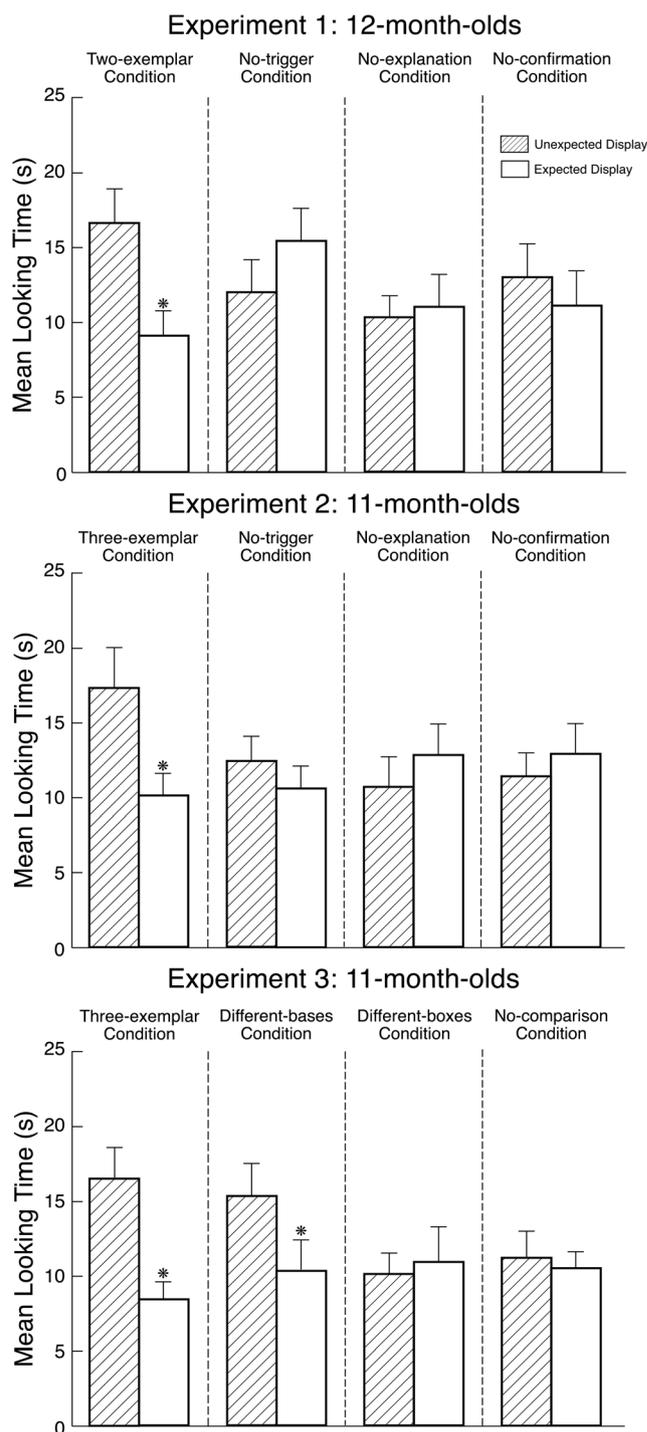


Fig. 3 Mean looking times at the unexpected and expected test displays in Experiments 1–3, separately for each condition. Errors bars represent standard errors, and an asterisk denotes a significant difference within a condition ($p < .05$ or better). Each condition had 20 infants. One additional group of 11-month-olds in the two-exemplar condition of Experiment 1 looked equally at the two displays (see the section Further Results With 11-Month-Olds).

arrive at an explanation for the small-on events and/or (b) more empirical evidence to confirm the new rule suggested by this explanation. As in Experiment 1, infants were assigned

to four conditions, and only the teaching trials differed across conditions.

Conditions

Three-exemplar condition Infants in the three-exemplar condition received three teaching pairs similar to those in the two-exemplar condition of Experiment 1; they saw the pink B-box and the green T-box in the first two teaching pairs, and a new, dark green staircase-shaped box (henceforth an S-box) in the third teaching pair (Fig. 4a). Across pairs, infants looked reliably longer at the small-on than at the large-on events, suggesting that they realized that the small-on events violated their proportion-of-contact rule. If three exemplars were sufficient for 11-month-olds to replace this rule with a more sophisticated proportional-distribution rule, then they should look reliably longer at the unexpected than at the expected test display.

No-trigger condition According to the EBL account, learning is triggered when infants encounter outcomes they cannot explain with their current rules. In the no-trigger condition of Experiment 1, all teaching events were consistent with infants' proportion-of-contact rule: The box remained stable when released with 50% of its bottom surface supported (large-on

events), and it fell when released with 0%–25% of its bottom surface supported (small-on events). In the no-trigger condition of Experiment 2, we explored a different approach: Infants saw no small-on events, only large-on events (Fig. 4b). The three large-on events from the three-exemplar condition were shown in two blocks of three trials. From a purely statistical standpoint, if infants detected that each asymmetrical box was always placed with its larger end on the base, then they should look reliably longer at the unexpected test display, which deviated from this regularity. According to the EBL account, however, because all teaching events were consistent with infants' proportion-of-contact rule, EBL should not be triggered, and infants should thus look equally at the two test displays.

No-explanation condition According to the EBL account, even when triggered by unexplained outcomes, the EBL process will come to a halt if infants are unable to build an explanation for these outcomes. As in Experiment 1, the teaching events in the no-explanation condition had reverse outcomes (Fig. 4c). Infants received the same three teaching pairs as in the three-exemplar condition, but the box now fell in the large-on events and remained stable in the small-on events. Analysis of the teaching trials indicated that infants looked reliably longer at the large-on than at the small-on events, suggesting that they

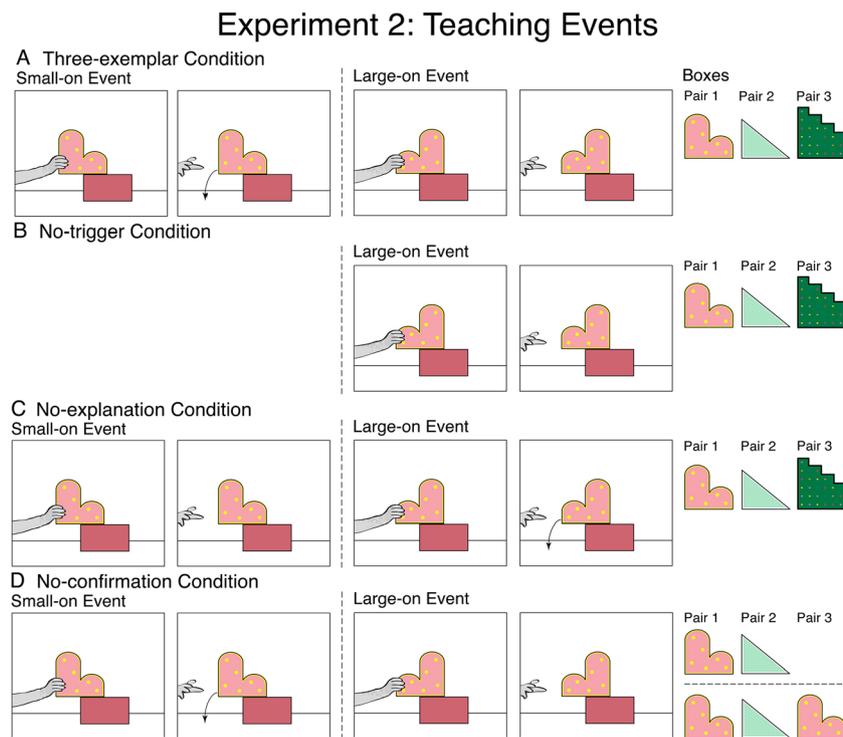


Fig. 4 Schematic depiction of the teaching events in each condition of Experiment 2. In the three-exemplar and no-explanation conditions, infants received three teaching pairs; the events from the first teaching pair are depicted, and the boxes used across pairs are depicted at the end of each row. In the no-confirmation condition, half of the infants received

only two teaching pairs, and half received a third teaching pair with the same box as in the first pair. In the no-trigger condition, infants saw three large-on events, with different boxes, in two identical blocks of three trials.

realized that the large-on events contradicted their proportion-of-contact rule. Nevertheless, the EBL account predicted that infants would be unable to generate an explanation for these events, and hence would look equally at the two test displays.

No-confirmation condition In the Further Results With 11-Month-Olds above, we reported that 11-month-olds tested as in the two-exemplar condition of Experiment 1 failed to learn the proportional-distribution rule, suggesting that these younger infants required (a) more information to arrive at an explanation for the small-on events and/or (b) more empirical evidence to confirm the rule suggested by this explanation. In the no-confirmation condition of Experiment 2, we sought to replicate this negative finding (Fig. 4d). Half of the infants again received two teaching pairs, with the pink B-box and green T-box; the other infants also received a third teaching pair with the pink B-box (this within-condition manipulation did not affect the test responses, suggesting that infants did require three distinct exemplars, rather than three teaching pairs, to learn the proportional-distribution rule).

Method

Participants

The participants were 80 healthy term 11-month-olds (39 male, 41 female; $M = 10;27$, range = 10;17–11;10), 20 per condition. Another 14 infants, distributed across conditions, were excluded because they were overly fussy, distracted (e.g., by their clothes), or active (ten), were inattentive during the teaching trials (three), or had a mean looking-time difference between the two test displays over three standard deviations from the condition mean (one).

Apparatus, stimuli, and procedure

The apparatus, stimuli, and procedure were identical to those in Experiment 1, with the addition (where specified above) of a third teaching pair. The dark green S-box ($27 \times 27 \times 8$ cm) had four steps and was decorated with small multicolored musical notes and outlined with black tape.

Results and discussion

Infants' test looking times (Fig. 3) were compared in an ANOVA with condition (three-exemplar, no-trigger, no-explanation, or no-confirmation) as a between-subjects factor and display (unexpected or expected) as a within-subjects factor. The analysis yielded only a significant Condition \times Display interaction, $F(3, 76) = 2.82$, $p = .045$, $\eta_p^2 = .10$. Planned comparisons revealed that the infants in the three-exemplar condition looked reliably longer at the unexpected ($M = 17.3$, $SD =$

12.2) than at the expected ($M = 10.1$, $SD = 6.7$) display, $F(1, 76) = 7.99$, $p = .006$, $d = 0.73$, whereas the infants in the no-trigger [unexpected: $M = 12.4$, $SD = 7.3$; expected: $M = 10.6$, $SD = 6.9$; $F(1, 76) < 1$, $d = 0.25$], no-explanation [unexpected: $M = 10.7$, $SD = 9.0$; expected: $M = 12.8$, $SD = 9.2$; $F(1, 76) < 1$, $d = -0.23$], and no-confirmation [unexpected: $M = 11.4$, $SD = 6.9$; expected: $M = 12.9$, $SD = 9.2$; $F(1, 76) < 1$, $d = -0.18$] conditions looked about equally at the two displays. Responses in the no-confirmation condition (with only two distinct exemplars in the teaching events) were reliably different from those of the 12-month-olds in the two-exemplar condition of Experiment 1, $F(1, 38) = 6.02$, $p = .019$, $\eta_p^2 = .14$, but were similar to those of the 11-month-olds in that same condition (Further Results With 11-Month-Olds), $F(1, 38) < 1$, $\eta_p^2 = .01$.

The results of Experiment 2 supported two conclusions. First, our results provided further evidence for the EBL account. In the three-exemplar condition, infants replaced their proportion-of-contact rule with a more sophisticated proportional-distribution rule, enabling them to detect the violation in the unexpected test display about two months before infants typically do so. In the remaining conditions, the teaching events failed to trigger or support the EBL process; as a result, infants continued to apply their proportion-of-contact rule, and hence failed to detect the violation in the unexpected test display. Second, unlike 12-month-olds, who needed only two exemplars to acquire the proportional-distribution rule, 11-month-olds required three exemplars. As we noted earlier, these younger infants may have required (a) more information to generate an explanation for the small-on events and/or (b) more empirical evidence to adopt the new rule suggested by this explanation.

Experiment 3

Experiment 3 had two goals: One was to confirm the positive results of the three-exemplar condition in Experiment 2, and the other was to begin exploring the explanation-building step in the EBL process. If explanations for support events are constructed via inferences from physical-domain knowledge, as the EBL account contends, then complicating this inference process should compromise learning.

In the three-exemplar condition of Experiment 2, two features of the teaching events might have helped infants arrive at an explanation for the small-on events: For each box, the only difference between the small-on and large-on events was the box's orientation, and the two events were shown in successive trials, making them easy to compare. In Experiment 3, we modified these two features. In two conditions, the large-on events differed from the small-on events not only in the box's orientation, but also in an additional, causally irrelevant way. In a third condition, the large-on events either were absent or

were presented in a separate block after the small-on events. These manipulations did not alter the explanation for the small-on events, but they did complicate the search for this explanation, because there was now more information for infants to consider and/or more demand on their limited working memory.

Conditions

Three-exemplar condition Infants received the same teaching pairs as in the three-exemplar condition of Experiment 2 (Fig. 5a). Across pairs, infants looked reliably longer at the small-on than at the large-on events, suggesting that they realized that the small-on events violated their proportion-of-contact rule. We predicted that, as in Experiment 2, infants would acquire the proportional-distribution rule and look reliably longer at the unexpected test display.

Different-bases condition Infants received the same teaching pairs as in the three-exemplar condition, except that a novel granite-gray base was used in the large-on events (Fig. 5b). Across teaching pairs, infants looked reliably longer at the small-on than at the large-on events, suggesting that they

realized that the small-on events violated their proportion-of-contact rule. It was unclear whether infants would still succeed in generating an explanation for these events, since there was now more outcome-predictive information for them to evaluate.

Different-boxes condition Infants received the same teaching pairs as in the three-exemplar condition, with one exception: In each pair, the box in the large-on event differed in color and pattern from that in the small-on event (Fig. 5c). To introduce the two B-, T-, and S-boxes, at the start of the session infants received three familiarization trials, one for each pair of boxes (in the order listed). In each trial, the two boxes stood side by side, in their correct orientations, with the small-on box on the left and the large-on box on the right. Following these trials, infants received the three teaching pairs. Across pairs, they looked reliably longer at the small-on than at the large-on events, suggesting that they realized that the small-on events violated their proportion-of-contact rule. The predictions for the test trials were the same as in the different-bases condition.

No-comparison condition As we mentioned above, infants in the three-exemplar condition could easily compare the

Experiment 3: Teaching Events

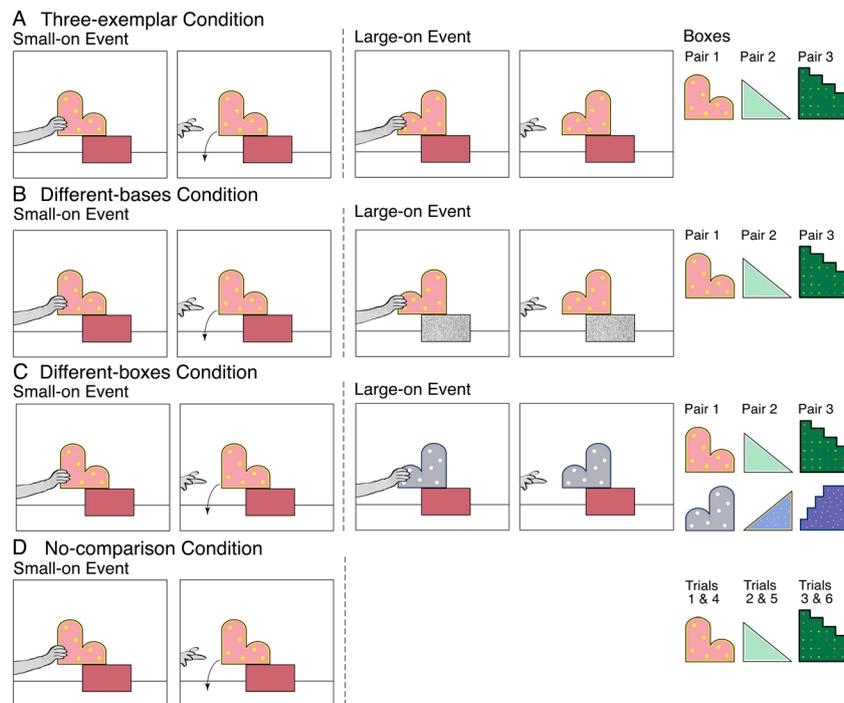


Fig. 5 Schematic depiction of the teaching events in each condition of Experiment 3. The three-exemplar condition was identical to that in Experiment 2. In the different-bases condition, a novel gray-granite base was used in all large-on events. In the different-boxes condition, a novel box (different in color and pattern) was used in the large-on event of each teaching pair. In the three-exemplar, different-bases, and different-boxes conditions, infants received three teaching pairs; the events from the first

teaching pair are depicted, and the boxes used across pairs are depicted at the end of each row. In the no-comparison condition, half of the infants saw the three small-on events from the three-exemplar condition in two identical blocks of three trials (as shown); the other infants saw the three large-on events from the three-exemplar condition in a second block.

small-on and large-on events for each box, since these events were always shown in successive trials. In the no-comparison condition, this easy comparison was no longer possible. Half of the infants saw no large-on events (Fig. 5d); instead, the small-on events from the three-exemplar condition were presented twice, in two identical blocks of three trials. The other infants saw the same small-on and large-on events as in the three-exemplar condition, but arranged in two blocks of three trials, beginning with the small-on events (this within-condition manipulation did not affect the test responses). If easy comparison of the small-on and large-on events helped the infants in the three-exemplar condition find the explanation for the small-on events, then infants in the no-comparison condition might fail to do so, and hence might look equally at the two test displays.

Method

Participants

The participants were 80 healthy term 11-month-olds (40 male, 40 female; $M = 11;1$, range = 10;18–11;14), 20 per condition. Another 15 infants, distributed across conditions, were excluded because they were overly fussy, distracted, or active (11); were inattentive during the teaching trials (two); or had a mean looking-time difference between the two test displays over three standard deviations from the condition mean (two).

Apparatus, stimuli, and procedure

The apparatus, stimuli, and procedure were identical to those in Experiment 2, with two exceptions. First, additional stimuli were used. The novel gray-granite base in the different-bases condition was otherwise identical to the red base, and the novel boxes in the different-boxes conditions included a light gray B-box decorated with large white dots and outlined with light blue tape, a light blue T-box decorated with small yellow dots and outlined with yellow tape, and a dark purplish-blue S-box decorated with small silver stars and outlined with blue tape. Second, each static familiarization trial in the different-boxes conditions ($M = 12.4$, $SD = 8.3$) ended when infants had (a) looked away for two consecutive seconds after having looked for at least four cumulative seconds or (b) looked for 60 cumulative seconds.

Results and discussion

Infants' test looking times (Fig. 3) were compared in an ANOVA with condition (three-exemplar, different-bases, different-boxes, or no-comparison) as a between-subjects factor and display (unexpected or expected) as a within-subjects factor. The analysis yielded a significant main effect of display,

$F(1, 76) = 8.06$, $p = .006$, and a significant Condition \times Display interaction, $F(3, 76) = 3.21$, $p = .028$, $\eta_p^2 = .11$. Planned comparisons revealed that infants in the three-exemplar condition looked reliably longer at the unexpected ($M = 16.5$, $SD = 9.1$) than at the expected ($M = 8.4$, $SD = 4.9$) display, $F(1, 76) = 12.66$, $p = .001$, $d = 1.09$; infants in the different-bases condition also looked reliably longer at the unexpected ($M = 15.3$, $SD = 9.5$) than at the expected ($M = 10.3$, $SD = 8.9$) display, $F(1, 76) = 4.80$, $p = .032$, $d = 0.54$; and infants in the different-boxes [unexpected: $M = 10.1$, $SD = 6.1$; expected: $M = 10.9$, $SD = 10.3$; $F(1, 76) < 1$, $d = -0.10$] and no-comparison [unexpected: $M = 11.2$, $SD = 7.7$; expected: $M = 10.5$, $SD = 4.6$; $F(1, 76) < 1$, $d = 0.11$] conditions looked about equally at the two displays.

The results of Experiment 3 supported two conclusions. First, the positive results of the three-exemplar and different-bases conditions confirmed those of Experiment 2 and provided further evidence for the EBL account. Second, the negative results in the different-boxes and no-comparison conditions made clear that, at this age and in this task, exposure to the small-on events alone was not sufficient for infants to acquire the proportional-distribution rule.

There are at least two ways in which the large-on events may have contributed to infants' success. First, these events demonstrated to infants that their proportion-of-contact rule sometimes applied in this novel laboratory situation; although their rule did not predict the outcomes of the small-on events, it did predict those of the large-on events. This partial failure/partial success may have encouraged infants to revise their rule. Second, seeing small-on and large-on events that were minimally different on successive trials may have helped infants rapidly zero in on the information needed to explain the unpredicted outcomes of the small-on events. In the three-exemplar condition, the small-on and large-on events differed only in the box's orientation; in the different-bases condition, the events also differed in their bases, but two factors may have minimized the impact of this additional change. One is that infants' physical-domain knowledge may have led them to swiftly discard the possibility that the color and pattern of the base could affect its ability to block the box's fall. The other factor is that the *same* novel base was used in all large-on events; once this change was deemed irrelevant, it could be ignored in subsequent teaching pairs, lessening the load on infants' working memory (this contrasted with the different-boxes condition, in which a *different* novel box was introduced in each large-on event).

General discussion

The present experiments focused on support events and attempted to help infants replace an early proportion-of-contact rule (that an object is stable when half or more of its

bottom surface is supported) with a more sophisticated proportional-distribution rule (that an object is stable when half or more of the entire object is supported); this rule is typically not acquired until about 13 months of age. Our experiments yielded four conclusions.

First, when shown teaching events that facilitated the EBL process, 11- and 12-month-olds acquired the rule: They subsequently detected a violation in an unexpected test display in which an L-shaped box remained stable with the right half of its bottom surface supported. Successful learning depended on exposure to two distinct exemplars at 12 months and three distinct exemplars at 11 months. Across the four successful conditions in Experiments 1–3, 59/80 infants (74%) looked longer at the unexpected than at the expected display, $p = .0000$ (cumulative binomial probability).

Second, when shown teaching events that failed to trigger or that disrupted the EBL process, infants did not acquire the rule. Thus, infants looked equally at the two test displays (a) when shown only teaching events consistent with their proportion-of-contact rule (*no-trigger* conditions); (b) when shown reverse teaching events for which they could construct no plausible explanation (*no-explanation* conditions); and (c) when shown too few distinct exemplars to confirm the rule (*no-confirmation* conditions). Only 69/140 infants (49%) in these conditions looked longer at the unexpected display, $p = .600$; this proportion differed reliably from that in the successful conditions, $p = .0004$ (Fisher exact test).

Third, infants also failed to acquire the rule when shown teaching events that could in principle support EBL but made the search for an explanation harder. Specifically, infants failed (a) when salient causally irrelevant differences were added to the teaching events that were consistent and inconsistent with the proportion-of-contact rule (*different-boxes* condition), and (b) when comparison of these events was made more difficult or impossible (*no-comparison* condition). Only 20/40 infants in these conditions looked longer at the unexpected display, $p = .563$; this proportion differed reliably from that in the successful conditions, $p = .014$ (Fisher exact test).

Finally, our results confirm previous findings that infants ages 11–12 months have not yet acquired the proportional-distribution rule. Leaving aside the no-explanation conditions, which showed reverse teaching events inconsistent with the rule and would have been confusing to infants who already knew the rule, all other unsuccessful conditions showed teaching events consistent with the rule. In these conditions, only 73/140 (52%) infants looked longer at the unexpected display, $p = .336$, suggesting that most infants had not yet acquired the rule.

Alternative interpretations

We have argued that EBL can account for our results. Could other learning mechanisms do so, as well? Below we consider two alternative possibilities.

Statistical learning First, consider any of the standard statistical-learning mechanisms, which have few constraints on what rules can be learned (e.g., Hastie, Tibsirani, & Friedman, 2009; Murphy, 2012). From a purely statistical perspective, it is difficult to explain why negative results were obtained in any of the conditions that presented regular patterns. For example, why did infants in the different-boxes and no-comparison conditions of Experiment 3 not learn that the box always fell when released with its smaller end on the base? Or why did infants in the no-trigger condition of Experiment 2 not learn that the box was always placed with its larger end on the base? It might be countered that the statistical patterns in these unsuccessful conditions were simply harder for infants to detect, for ancillary reasons having to do with perceptual salience, working-memory limitations, and so on.

This could not be the case for the no-explanation conditions of Experiments 1 and 2, however: Apart from their reverse outcomes, these conditions were identical to the successful conditions. Why, then, did infants fail to learn the reverse pattern they were shown? One suggestion might be that (a) many pertinent observations are necessary for infants to learn the rather complex proportional-distribution rule using statistics alone; (b) the infants in our experiments had begun accumulating such observations in daily life and required only three or fewer observations to finally learn the rule; and (c) the infants in the no-explanation conditions were confused when shown reverse outcomes that conflicted with their stored observations.

Although this suggestion offers an explanation for the results of the no-explanation conditions, it cannot explain those of the successful conditions. To see why, suppose that our 11- and 12-month-olds were indeed in the midst of statistically learning the proportional-distribution rule from many observations collected over weeks or months. At the time of their participation, the infants would have fallen into one of three groups: (a) those who had already learned the rule; (b) those who had not yet learned the rule but needed only three or fewer observations to do so; and (c) those who had not yet learned the rule and needed more than three observations to do so. If the infants were using standard statistical learning, we would expect group (b) to be small, with most infants being in group (c) and perhaps a few in group (a). In fact, our results painted a different picture: group (b) was large (i.e., the majority of 11- and 12-month-olds in the successful conditions learned the rule), whereas groups (a) and (c) were small.

Hierarchical Bayesian learning A hierarchical Bayesian learner could easily be designed to acquire the proportional-distribution rule with only two to three observations, as in the successful conditions. However, the learner would do so in a very different way from EBL.

A Bayesian model is a general method for describing complex world interactions (e.g., Darwiche, 2009; A. Gelman et al., 2013; Koller & Friedman, 2009; Lee, 2011; Leonard & Hsu, 1999; Perfors, Tenenbaum, Griffiths, & Xu, 2011). It consists of three parts: a set of world features, a specification of which features directly influence each other, and parameters governing these influences. The first two parts are typically captured by a graph of nodes (the features) and links (the direct influences). The parameters govern local interactions among the directly connected features, and distal interactions are inferred by propagating information through the network. Generally, the designer of a Bayesian model provides a graphical structure and a subjective prior distribution of initial parameter values. Together, these make predictions possible: When a particular feature is observed, the model can predict its effects on other features of the world. “Learning” in the Bayesian framework typically refers to adjusting the parameter values to fit observations of the world, thereby transforming the prior distribution into a posterior distribution. In a hierarchical Bayesian learner, higher-level latent (i.e., not directly observable) features can exert a systematic influence over lower-level features. For example, whether dice are loaded and whether an individual is honest are useful higher-level features; although they cannot be directly observed, they influence lower-level features and can be helpful in guiding predictions. The parameters that govern the influences of higher-level features are termed *hyperparameters*.

To model our successful conditions, a hierarchical Bayesian learner would employ a set of lower-level features describing the box, the base, and so on, as well as two higher-level latent features: a *proportion-of-contact* feature and a fledgling alternative feature that would become *proportional distribution*. Importantly, the latter feature would have to already be present in the network in some rudimentary form and be properly connected to other features, although its hyperparameters might be very approximate. A strong prior would be provided for the *proportion-of-contact* feature, reflecting the learner’s existing familiarity with this feature and its influence over lower-level features; a weak prior would be provided for the fledgling *proportional-distribution* feature and its effects. The seemingly anomalous observations provided during training (i.e., the box fell even though half of its bottom surface was supported by the base) would cause the learner to ascribe the box’s behavior to the *proportional-distribution* feature; adjusting the relatively weaker, and hence more malleable, hyperparameters associated with this feature would be preferred over adjusting the strong hyperparameters associated with the *proportion-of-contact* feature. This parametric adjustment would allow the learner to acquire the *proportional-distribution* rule with only a few observations, as in our successful conditions.

In contrast to the parametric learning outlined above, structural learning is problematic in Bayesian models, except in special cases (see, e.g., Chow & Liu, 1968; Loh &

Wainwright, 2013; Oates, Smith, & Mukherjee 2016; Rebane & Pearl, 1987); this is due to the fact that conditional independence, the foundation of the Bayesian models’ effectiveness and utility, is an analytic statistical property and not an empirical one. If one knew which new node to add to a graphic structure and how to connect it to existing nodes, it would then be easy to verify that this structural change does improve the model’s performance. But selecting which structural change to make is computationally intractable, for two reasons. First, selecting the right change (like selecting the winning lottery ticket) is virtually impossible, because there are far too many alternatives to choose from. Second, because the space of possible Bayesian models is highly nonconvex, evaluating one possible structural change in general gives little information about how others will fare.

Structural learning in EBL

In contrast to a Bayesian learner, an EBL learner is able to add new features to its world representation. To explain the responses of infants in our successful conditions, EBL does not require the prior existence of a rudimentary *proportional-distribution* feature. Rather, that feature is introduced by inference over imperfect but general physical-domain knowledge, which includes core knowledge and previously acquired rules. The result is new general physical-domain knowledge that, although still imperfect, represents a significant improvement over the original knowledge. From an EBL point of view, core knowledge represents general patterns of world behavior that, over evolutionary periods, have found their way into our DNA. Apart from its initial bootstrapping function, however, core knowledge occupies no special status in EBL; it can be imperfect or approximate, and it may be eclipsed as more accurate general knowledge is acquired.

In this final section, we briefly consider how an artificial-intelligence EBL system might demonstrate structural learning in response to the observations in the three-exemplar conditions of Experiments 2 and 3. Our goal here is not to model infants’ knowledge and reasoning, but rather to offer an algorithmic existence proof of structural learning in an EBL system, by outlining the kind of processing that occurs.

Let us suppose that the system’s initial domain knowledge includes the following set of core and previously acquired rules (numbered 1–4 for convenience only):

1. An object that is not supported falls.
2. An object on a base is adequately supported if half or more of its bottom surface rests on the base.
3. An object behaves as a unit.
4. Larger effects overwhelm smaller ones.

When shown the first small-on teaching event (with the B-box), the system detects that the box’s behavior contradicts Rule

2: The box falls even though the right half of its bottom surface rests on the base. This unexplained observation triggers the search for an explanation. Sooner or later, the system entertains the possibility of viewing the box as two connected subobjects; by mentally decomposing the box in a specific way, the box's behavior can be explained by chaining together existing rules. Specifically, if the box is cut vertically from the left edge of the base, left and right subobjects are formed, and the expected behavior of each subobject becomes clear: By Rule 1, the fully unsupported left subobject must fall, and by Rule 2, the fully supported right subobject must not fall. However, by Rule 3, the imaginary cut cannot result in incompatible outcomes. Using Rule 4, the effect of the larger left subobject wins out and, as observed, the entire box falls. This explanation allows the EBL system to conjecture a new rule: When an object is released with one end on a base, it will fall if the proportion of the entire object off the base is greater than that on the base. In this way, a conceptual feature that was not present previously makes possible the new proportional-distribution rule

This candidate rule must then be confirmed empirically. Because the derivation process provides significant analytic evidence for the rule, however, only a few additional observations (with the T-box and the S-box) suffice to ensure that the explanation for the original observation was not specious. Once confirmed, the conjectured rule (name it Rule 5) is added to the domain knowledge. It will then remain in the rule set or be discarded, depending on whether its benefit (the improved prediction of world behavior) outweighs its cost (the resources consumed in maintaining and entertaining it; e.g., Gratch & DeJong, 1996; Greiner & Jurisica, 1992; Minton, Carbonell, Etzioni, Knoblock, & Kuokka, 1987). If it is kept, it will be available to participate in further explanations, so that more sophisticated world interactions may become explainable. Finally, the addition of Rule 5 alters the utility of other rules; for example, Rule 2 will likely be discarded as no longer achieving a positive cost/benefit.

This algorithmic account has two implications for our discussion of the EBL process in our successful conditions. First, if in their quest for an explanation infants mentally decomposed each asymmetrical teaching box into left and right portions, then in future research one could either support this process with congruent perceptual cues (e.g., coloring the left and right portions of each box differently) or interfere with it with incongruent cues (e.g., coloring the top and bottom portions of each box differently). Second, when considering the many different factors that can affect infants' ability to generate an explanation (including the factors identified in the different-boxes and no-comparison conditions of Exp. 3), it becomes clear why infants would be able to acquire the proportional-distribution rule earlier in a laboratory setting. The natural world will present infants with many small-on and large-on events from which to acquire the proportional-distribution rule via EBL. In the laboratory, however, confounds and possibilities for alternative explanations can be reduced to a

minimum, making the relevant explanation much easier for infants to discover.

Conclusions

In the present research, 11- and 12-month-olds learned a new support rule with very few observations when shown teaching events designed to facilitate EBL. Conversely, they failed to learn the rule when shown teaching events that derailed EBL. Together, these results demonstrate that despite their limited knowledge about the world, infants can still leverage this knowledge to benefit from EBL, making possible highly efficient learning.

Author note This research was supported by a grant from the NICHD (HD-21104) to R.B. We thank Frank Keil and Alan Leslie for helpful suggestions; Stephanie Sloane and the research staff at the UIUC Infant Cognition Laboratory for their help with the data collection; and the parents and infants who participated in the research.

References

- Baillargeon, R. (1995). A model of physical reasoning in infancy. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 9, pp. 305–371). Norwood: Ablex.
- Baillargeon, R. (1998). Infants' understanding of the physical world. In M. Sabourin, F. Craik, & M. Robert (Eds.), *Advances in psychological science* (Vol. 2, pp. 503–529). London: Psychology Press.
- Baillargeon, R. (1999). Young infants' expectations about hidden objects: A reply to three challenges. *Developmental Science*, 2, 115–132. doi:10.1111/1467-7687.00061
- Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, 3, 2–13. doi:10.1111/j.1745-6916.2008.00056.x
- Baillargeon, R., & Carey, S. (2012). Core cognition and beyond: The acquisition of physical and numerical knowledge. In S. Pauen (Ed.), *Early childhood development and later outcome* (pp. 33–65). Cambridge: Cambridge University Press.
- Baillargeon, R., & DeVos, J. (1991). Object permanence in 3.5- and 4.5-month-old infants: Further evidence. *Child Development*, 62, 1227–1246. doi:10.2307/1130803
- Baillargeon, R., Li, J., Gertner, Y., & Wu, D. (2011). How do infants reason about physical events? In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 11–48). Oxford: Blackwell.
- Baillargeon, R., Li, J., Ng, W., & Yuan, S. (2009). An account of infants' physical reasoning. In A. Woodward & A. Needham (Eds.), *Learning and the infant mind* (pp. 66–116). New York: Oxford University Press.
- Baillargeon, R., Needham, A., & DeVos, J. (1992). The development of young infants' intuitions about support. *Early Development and Parenting*, 1, 69–78. doi:10.1002/edp.2430010203
- Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in 5-month-old infants. *Cognition*, 20, 191–208. doi:10.1016/0010-0277(85)90008-3
- Baillargeon, R., Stavans, M., Wu, D., Gertner, Y., Setoh, P., Kittredge, A. K., & Bernard, A. (2012). Object individuation and physical reasoning in infancy: An integrative account. *Language Learning and Development*, 8, 4–46. doi:10.1080/15475441.2012.630610

- Baillargeon, R., Wu, D., Yuan, S., & Luo, Y. (2009). Young infants' expectations about self-propelled objects. In B. Hood & L. Santos (Eds.), *The origins of object knowledge* (pp. 285–352). Oxford: Oxford University Press.
- Bishop, C. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.
- Chow, C. K., & Liu, C. N. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14, 462–467.
- Dan, N., Omori, T., & Tomiyasu, Y. (2000). Development of infants' intuitions about support relations: Sensitivity to stability. *Developmental Science*, 3, 171–180. doi:10.1111/1467-7687.00110
- Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks*. New York: Cambridge University Press.
- DeJong, G. F. (Ed.). (1993). *Investigating explanation-based learning*. Boston: Kluwer Academic Press.
- DeJong, G. F. (2014). Explanation-based learning. In T. Gonzalez, J. Diaz-Herrera, & A. Tucker (Eds.), *CRC computing handbook: Computer science and software engineering* (3rd ed., pp. 66.1–66.26). Boca Raton: CRC Press.
- Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, 14, 79–106. doi:10.1207/s15516709cog1401_5
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton: CRC Press.
- Gratch, J., & DeJong, G. F. (1996). A decision-theoretic approach to adaptive problem solving. *Artificial Intelligence*, 88, 101–142.
- Greiner, R., & Jurisica, I. (1992). A statistical approach to solving the EBL utility problem. *Proceedings of the Tenth National Conference on Artificial Intelligence* (San Jose, CA), pp. 241–248.
- Hastie, T., Tibsiran, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York: Springer.
- Hespos, S. J., & Baillargeon, R. (2001). Knowledge about containment events in very young infants. *Cognition*, 78, 207–245. doi:10.1016/S0010-0277(00)00118-9
- Hespos, S. J., & Baillargeon, R. (2006). Décalage in infants' knowledge about occlusion and containment events: Converging evidence from action tasks. *Cognition*, 99, B31–B41. doi:10.1016/j.cognition.2005.01.010
- Hespos, S. J., & Baillargeon, R. (2008). Young infants' actions reveal their developing knowledge of support variables: Converging evidence for violation-of-expectation findings. *Cognition*, 107, 304–316. doi:10.1016/j.cognition.2007.07.009
- Huettel, S. A., & Needham, A. (2000). Effects of balance relations between objects on infants' object segregation. *Developmental Science*, 3, 415–427. doi:10.1111/1467-7687.00136
- Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 234–262). Oxford: Oxford University Press, Clarendon Press.
- Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. Cambridge: MIT Press.
- Lee, M. (Ed.). (2011). Hierarchical Bayesian models (Special issue). *Journal of Mathematical Psychology*, 55(1).
- Leonard, T., & Hsu, J. (1999). *Bayesian methods*. Cambridge: Cambridge University Press.
- Leslie, A. M. (1995). A theory of agency. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 121–149). Oxford: Oxford University Press, Clarendon Press.
- Loh, P., & Wainwright, M. J. (2013). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *Annals of Statistics*, 41, 3022–3049.
- Luo, Y., & Baillargeon, R. (2005). When the ordinary seems unexpected: Evidence for incremental physical knowledge in young infants. *Cognition*, 95, 297–328. doi:10.1016/j.cognition.2004.01.010
- Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology*, 58, 441–486. doi:10.1016/j.cogpsych.2008.11.001
- Minton, S., Carbonell, J. G., Etzioni, O., Knoblock, C. A., & Kuokka, D. R. (1987). Acquiring effective search control rules: Explanation-based learning in the PRODIGY system. In P. Langley (Ed.), *Proceedings of the Fourth International Workshop on Machine Learning* (pp. 122–133). Amsterdam: Elsevier.
- Mitchell, T. (1997). *Machine learning*. New York: McGraw Hill.
- Murphy, K. (2012). *Machine learning: A probabilistic perspective*. Cambridge: MIT Press.
- Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, 47, 121–148. doi:10.1016/0010-0277(93)90002-D
- Needham, A., & Baillargeon, R. (1997). Object segregation in 8-month-old infants. *Cognition*, 62, 121–149. doi:10.1016/S0010-0277(96)00727-5
- Oates, C., Smith, J., & Mukherjee, S. (2016). Estimating causal structure using conditional DAG models. *Journal of Machine Learning Research*, 17, 1–23.
- Perfors, A., Tenenbaum, J., Griffiths, T., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, 120, 302–321. doi:10.1016/j.cognition.2010.11.015
- Rebane, G., & Pearl, J. (1987). The recovery of causal poly-trees from statistical data. In *Proceedings of the 3rd Workshop on Uncertainty in Artificial Intelligence* (pp. 222–228). Arlington: AUAI Press.
- Setoh, P., Wu, D., Baillargeon, R., & Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences*, 110, 15937–15942. doi:10.1073/pnas.1314075110
- Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, 8, 481–520. doi:10.1016/0010-0285(76)90016-5
- Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology*, 36, 273–310. doi:10.1006/cogp.1998.0686
- Spelke, E. S. (1994). Initial knowledge: Six suggestions. *Cognition*, 50, 431–445. doi:10.1016/0010-0277(94)90039-6
- Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 99, 605–632. doi:10.1037/0033-295X.99.4.605
- Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). Oxford: Oxford University Press, Clarendon Press.
- Wang, S., & Baillargeon, R. (2006). Infants' physical knowledge affects their change detection. *Developmental Science*, 9, 173–181. doi:10.1111/j.1467-7687.2006.00477.x
- Wang, S., & Baillargeon, R. (2008). Can infants be “taught” to attend to a new physical variable in an event category? The case of height in covering events. *Cognitive Psychology*, 56, 284–326. doi:10.1016/j.cogpsych.2007.06.003
- Wang, S., Baillargeon, R., & Paterson, S. (2005). Detecting continuity violations in infancy: A new account and new evidence from covering and tube events. *Cognition*, 95, 129–173. doi:10.1016/j.cognition.2002.11.001
- Wang, S., & Kohne, L. (2007). Visual experience enhances 9-month-old infants' use of task-relevant information in an action task. *Developmental Psychology*, 43, 1513–1522. doi:10.1037/0012-1649.43.6.1513
- Wang, S., Zhang, Y., & Baillargeon, R. (2016). Young infants view physically possible support events as unexpected: New evidence for rule learning. *Cognition*, 157, 100–105. doi:10.1016/j.cognition.2016.08.021