The Violation-of-Expectation Paradigm: A Conceptual Overview

Francesco Margoni[a,b], Luca Surian[c], and Renée Baillargeon[d]

[a] Department of Social Studies, University of Stavanger, Norway

[b] Department of Psychology, University of Oslo, Norway

[c] Department of Psychology and Cognitive Science, University of Trento, Italy

[d] Department of Psychology, University of Illinois at Urbana-Champaign, USA

Authors' Notes

Francesco Margoni ORCID ID 0000-0002-8109-228X

Luca Surian ORCID ID 0000-0001-6463-2953

Renée Baillargeon ORCID ID 0000-0002-1177-9940

Correspondence should be addressed to: Francesco Margoni, University of Stavanger, Institutt for Sosialfag, 4021 Stavanger, Norway, email: francesco.margoni@uis.no, or Renée Baillargeon, Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820, USA, email: rbaillar@illinois.edu.

Citation

Margoni, F., Surian, L., & Baillargeon, R. (2024). The violation-of-expectation paradigm: A conceptual overview. *Psychological Review*, *131*(3), 716-748.

https://doi.org/10.1037/rev0000450.

This article was accepted for publication in *Psychological Review* on Aug 14 2023.

Abstract

For over 35 years, the violation-of-expectation paradigm has been used to study the development of expectations in the first three years of life. A wide range of expectations has been examined, including physical, psychological, sociomoral, biological, numerical, statistical, probabilistic, and linguistic expectations. Surprisingly, despite the paradigm's widespread use and the many seminal findings it has contributed to psychological science, so far no one has tried to provide a detailed and in-depth conceptual overview of the paradigm. Here, we attempted to do just that. We first focus on the rationale of the paradigm and discuss how it has evolved over time. We then show how improved descriptions of infants' looking behavior, together with the addition of a rich panoply of brain and behavioral measures, have helped deepen our understanding of infants' responses to violations. Next, we review the paradigm's strengths and limitations. Finally, we end with a discussion of challenges that have been leveled against the paradigm over the years. Through it all, our goal was two-fold. First, we sought to provide psychologists and other scientists interested in the paradigm with an informed and constructive analysis of its theoretical origins and development. Second, we wanted to take stock of what the paradigm has revealed to date about how infants reason about events, and about how surprise at unexpected events, in or out of the laboratory, can lead to learning, by prompting infants to revise their working model of the world.

Keywords: cognitive development, infant cognition, measures of surprise, expectation, explanation.

1. Introduction

For over three decades, the violation-of-expectation (VOE) paradigm has been used widely to study many different facets of infant cognition. It has produced numerous findings that have become part of the empirical basis of developmental psychology, and it has contributed substantially to theoretical advances in that discipline. Surprisingly, no one has yet put forth a detailed and in-depth conceptual overview of the VOE paradigm. Here, we aimed to do just that: reflect on how the paradigm has evolved from its first emergence to its many present-day extensions; probe its strengths and limitations; and consider its challenges. Through it all, we had two main goals in mind. One was to provide psychologists and other scientists interested in the paradigm with an informed and constructive analysis of its theoretical origins and development. The other was to consider what VOE findings have revealed (a) about how infants represent, reason about, and respond to events, and (b) about how surprise at unexpected events, in or out of the laboratory, can lead to learning, by prompting infants to revise their working model of the world.

To prevent misconceptions about our article, we begin with three disclaimers. First, we emphasize that our article in no way offers an exhaustive review of the myriad of empirical findings that have been obtained with the VOE paradigm over the past three decades. In each section of the article, we describe only a few findings, to illustrate and support general points about the paradigm. In sections 2 and 3, we focus on VOE tasks that measured infants' total looking time, historically the first measure used in such tasks. In sections 4 and 5, we discuss VOE tasks that introduced a rich panoply of other measures and helped deepen our understanding of how infants represent events, reason about them, and respond to violations.

Second, we acknowledge that our review is not theory-neutral: We take positive evidence that infants respond differently to events that violate as opposed to accord with an expectation, together with appropriate controls, to suggest that infants possess this expectation. We refer to such interpretations as *conceptual* interpretations: They grant infants expectations about the world as well as the capacity to bring to bear relevant expectations when interpreting events and reasoning about their outcomes. Conceptual interpretations stand in contrast to *non-conceptual* interpretations, which do not grant infants expectations but invoke other, typically lower-level processes (e.g., familiarity preferences, perceptual biases, or domain-general memory processes). In sections 9.1 and 9.2, we consider a number of non-conceptual interpretations that have been proposed for particular VOE findings and argue that they do not adequately explain these findings. This is not meant to suggest that non-conceptual interpretations generally need not be considered when interpreting VOE findings. Because non-conceptual processes can also drive infants' responses to events, positive evidence that infants detect the violation in a VOE task cannot support the conclusion that they possess the expectation under investigation without the careful evaluation of plausible alternative non-conceptual interpretations. As we argue in section 8.1, not only must non-conceptual interpretations of VOE findings be routinely evaluated, but doing so often yields further insights into infants' expectations.

Third, we recognize that some VOE findings have given rise to divergent conceptual interpretations, some granting infants rich and abstract expectations and others leaner and more circumscribed expectations. Such disagreements have engendered vibrant debates in the infancy literature about what specific expectations are driving infants' responses (e.g., in the area of numerical reasoning; Carey, 2009; Leslie et al., 2008). Likewise, granting infants of a particular age expectations based on VOE findings typically says little about the mechanisms by which these expectations have been attained, giving rise to ongoing debates about the developmental origins of infants’ expectations (e.g., in the area of sociomoral reasoning; Buyukozer Dawkins et al., 2019; Ziv & Sommerville, 2017). We do not review such debates here but maintain our focus on the VOE paradigm and how it has been used to explore infants’ minds and uncover their expectations.

2. Conceptual Rationale and its Historical Development

Over the course of its history, the VOE paradigm has undergone substantial changes in its conceptual rationale. In this section, we review key phases in this evolution.

2.1. Beginnings

For a large part of the 20th century, Piaget’s (1952, 1954) stage theory dominated research on early cognitive development. According to this theory, young infants are limited sensorimotor processors incapable of representation and thought; nevertheless, as they begin to act on objects and learn from these interactions, they become capable of increasingly intelligent actions. Evidence for these conclusions came in part from Piaget’s research on the development of object permanence, the belief that objects continue to exist when out of view. Piaget observed that infants under 8 to 9 months of age typically fail to search for objects they have watched being hidden, and he concluded that infants are initially incapable of representing the continued existence of hidden objects. In time, however, investigators began to question this conclusion. Bower (1974), in particular, suggested that young infants might fail to search for hidden objects due to motor limitations alone, and he argued that non-search tasks were needed to determine whether young infants truly lacked object permanence.

During the 1960s and 1970s, Bower devised various non-search tasks to assess object permanence in young infants (Bower, 1974; Bower et al., 1971). Many of these tasks involved visual tracking: For example, an object moved back and forth along a track whose center was occluded by a large screen, and Bower measured whether infants anticipated the reappearance of the object after it moved behind the screen, or whether their tracking was disrupted if a different object reappeared. Based on the results of these investigations, Bower concluded that infants demonstrated object permanence by at least 5 months of age, contradicting Piaget's (1952, 1954) claims. As might be expected, this conclusion attracted a great deal of experimental attention, which overwhelmingly cast doubt on Bower's results. In particular, researchers pointed out that the results suffered from methodological confounds, were open to alternative interpretations consistent with Piagetian theory, or failed to replicate altogether (Meicler & Gratch, 1980; Muller & Aslin, 1978).

Although Bower's (1974; Bower et al., 1971) findings did not withstand experimental scrutiny, his argument that non-search tests of object permanence were needed to fully evaluate Piaget's (1952, 1954) claims was still sound. In their quest for such a test, Baillargeon et al. (1985) turned to a commonly used and well-accepted looking-time paradigm in infancy research, the habituation paradigm (Cohen, 1976; Fantz, 1964). In a typical task, infants are shown one stimulus repeatedly until their looking time decreases across trials to a pre-set habituation criterion. Next, on alternative test trials, infants are presented with the now familiar stimulus and a novel stimulus. The rationale is that if infants can discriminate between the two stimuli, they will look longer at the novel than at the familiar stimulus. By the 1980s, habituation experiments had brought to light many perceptual and cognitive abilities in young infants (Bornstein, 1985; Spelke, 1985). Baillargeon et al. speculated that if young infants looked longer not only at novel vs. familiar events but also at surprising, impossible events that violated object permanence vs. possible events that accorded with it, then this extension of the habituation paradigm might provide a less contentious approach to evaluating Piaget's claims about the development of object permanence.

To be clear, the innovation in this approach did not lie in presenting infants with impossible events that violated object permanence. Both Bower and his detractors had used such events in their experiments (e.g., in a visual-tracking task, an object might disappear behind one screen and reappear from behind another screen without appearing in the gap between them; Moore et al., 1978). Moreover, Charlesworth (1969) had long advocated the use of surprising, impossible events to study infants' object concept (e.g., in a search task, an object hidden under a cloth might no longer be there when the infant lifted the cloth). What was new in the approach of Baillargeon et al. (1985) was the reliance on looking times (as opposed to disruptions in visual tracking, facial expressions of surprise, etc.) to assess infants' responses to permanence violations.

2.2. Possible and Impossible Events

Baillargeon et al. (1985) first habituated 5.5-month-olds to a screen that rotated in depth through a 180^o arc, in the manner of a drawbridge. Following habituation, a box was placed in the path of the screen, and infants saw two test events. In the impossible event, the screen again rotated through a full 180^o arc, as though no box blocked its path. In the possible event, the screen rotated only through a 120^o arc, until it reached the hidden box. Infants looked significantly longer at the impossible than at the possible event, suggesting that they (a) represented the continued existence of the box after it became hidden by the screen, (b) expected the screen to stop when it reached the box, and (c) were surprised that it did not. This effect was replicated with 4.5-month-olds and with 3.5-month-old fast habituators (Baillargeon, 1987); it was eliminated when the box was placed out of the path of the screen (Baillargeon et al., 1985) or was removed altogether (Baillargeon, 1987). These results suggested that 3.5- to 5.5-month-olds already hold a notion of object permanence, casting doubts on Piaget's (1952, 1954) claims.

In the decade that followed (roughly 1985-1994), VOE reports using a wide array of impossible and possible events provided converging evidence that young infants represent the continued existence of hidden objects (Baillargeon, 1986; Spelke et al., 1992). Encouraged by these efforts, researchers began studying other facets of infants' understanding of the physical world, devising in each case possible and impossible events suitable for the purpose. These experiments revealed additional competencies in infants' ability to reason about the number and properties of objects in occlusion, support, collision, and other events (Baillargeon & Graber, 1987; Needham & Baillargeon, 1993; Woodward et al., 1993; Wynn, 1992). Alongside these advances in our understanding of early physical reasoning came new insights into the VOE paradigm responsible for them. In particular, researchers came to realize that although this paradigm had initially been construed as an extension of the habituation paradigm, it differed from it in important respects; for ease of discussion, we will refer to standard habituation tasks as *familiar-novel* tasks, and to habituation tasks such as the VOE rotating-screen task described above as *possible-impossible* tasks. From the start, there had been an awareness that the two types of tasks depend on different tendencies in infants. On the one hand, familiar-novel tasks depend on infants' natural tendency to look longer at novel than at familiar stimuli; this penchant for novelty motivates exploration and learning as infants are drawn to inspect new objects and events and discover their properties. On the other hand, possible-impossible tasks depend on infants' natural tendency to bring to bear their physical knowledge to form expectations about events' outcomes and to search for explanations when these expectations are violated; this predisposition for making sense of the world as it unfolds also drives learning and results in a more predictable world, with respect to both observation and prospective action. In time, it became clearer that these two different rationales—novelty and expectation violation—meant that habituation trials played different roles in the two types of tasks.

In a typical familiar-novel task involving two test events, event-A and event-B, neither event, when presented alone, draws more attention than the other. Following habituation to one event (e.g., event-A), however, the other event (event-B) becomes relatively novel and, as such, elicits more attention (assuming infants can distinguish the events). Habituation trials are thus essential for making one event familiar; without these trials, infants have no basis for responding differentially to the two events. In a possible-impossible task, however, habituation trials do not play the role of making one event more familiar than the other. As time went on, researchers using possible-impossible tasks began to vary both the number and the nature of the trials administered prior to test (henceforth *pretest* trials). Below, we describe three types of pretest trials, all of which are still in use today.

Introduction trials provide an introduction to subcomponents of the test events, to help infants process these events and zero in on the violation embedded in the impossible event. Both age and event complexity contribute to the number of introduction trials administered. With complex test events, infants may be fully habituated to an introduction event, as we saw in the rotating-screen experiment described earlier: Infants were fully habituated to the 180°-rotation of the screen prior to test (Baillargeon, 1987; Baillargeon et al., 1985). With simpler test events, however, infants may receive only a few introduction trials (Baillargeon & Graber, 1987), or even none at all (Needham & Baillargeon, 1993). For example, in an experiment with test trials only (Needham & Baillargeon, 1993), 4.5-month-olds saw test events in which an experimenter's gloved hand released a box either on a platform (possible event) or in midair next to the platform (impossible event); in each case, the box remained stationary when released.

Infants looked significantly longer at the impossible than at the possible event, suggesting that they expected the box to fall when released in midair and were surprised that it did not. This effect was eliminated when the gloved hand placed the box in the same positions but did not release it, thereby providing continuous support for it.

Modulation trials provide information that can render one of the two test events physically impossible; without that information, both events are equally possible. For example, in an experiment using a between-subjects design (Kotovsky & Baillargeon, 1994), 11-month-olds saw a single test event in which a large (large-cylinder event) or a small (small-cylinder event) cylinder rolled down a ramp and hit a wheeled toy bug on a track, causing it to roll to the end of the track, on the far side of the apparatus. Prior to the test trial, infants received six modulation trials involving a medium cylinder (infants manipulated the small, medium, and large cylinders before the testing session and were thus aware of their different weights). When the medium cylinder caused the bug to roll only to the *middle* of the track, infants looked significantly longer if shown the small-cylinder as opposed to the large-cylinder event, suggesting that (a) they expected the small cylinder to cause the bug to travel less far compared to the medium cylinder, and (b) they were surprised when this expectation was violated. When the medium cylinder caused the bug to roll to the *end* of the track, however, infants looked equally at the two test events, as they then had no unambiguous basis for calibrating their expectations (i.e., perhaps all three cylinders could cause the bug to travel to the end of the track).

Finally, *hint* trials provide a hint as to how the impossible test event was produced. These trials are typically used in control conditions; the rationale is that if infants can use the hint provided to generate an explanation for the impossible event, then they should no longer find that event surprising. To illustrate, in one experiment (Baillargeon & Graber, 1987), 5.5-month-olds first received four introduction trials in which a toy rabbit moved back and forth behind a large screen; in two trials, the rabbit was about as tall as the screen (tall-rabbit event), and in two trials the rabbit was half as tall (short-rabbit event). The impossible and possible test events were identical to the tall-rabbit and short-rabbit events, respectively, except that a window was now present in the upper half of the screen; neither rabbit appeared in the window as it moved back and forth behind the screen. Infants looked significantly longer at the impossible than at the possible event, suggesting that they expected the tall rabbit to appear in the window and were surprised that it did not. In a control condition, infants first received two static hint trials in which they saw two tall rabbits (tall-rabbit hint trial) or two short rabbits (short-rabbit hint trial) standing on either side of the screen. Infants now looked equally at the impossible and possible events, suggesting that they took advantage of the hint provided to generate an explanation for the impossible event: Two tall rabbits moved behind the screen, one on each side of the window.

Although for the sake of clarity we have described introduction, modulation, and hint trials as serving distinct functions, in practice the same pretest trials can serve more than one function. For example, modulation trials can serve both to introduce subcomponents of the test events and to provide critical information that renders one of the events impossible.

In sum, by the end of this first decade of VOE research (1985-1994), many different possible-impossible reports had been published that varied the number and type of pretest trials administered. This also had implications for how the experimental and control conditions differed in each task. In some tasks, different test events were shown in the two conditions (Needham & Baillargeon, 1993). In other tasks, the same test events were shown in the two conditions, but modulation trials in the experimental condition led infants to perceive one of the events as impossible, eliciting surprise at the event (Kotovsky & Baillargeon, 1994). In yet other tasks, the same impossible and possible events were shown in the two conditions, but hint trials in the control condition suggested how the impossible event was produced, thereby eliminating infants' surprise at the event (Baillargeon & Graber, 1987).

All of these variations highlighted the differences between familiar-novel and possible-impossible tasks. As a result, researchers began to refer to the VOE paradigm as separate and distinct from the habituation paradigm. At the time, a VOE experiment was defined as one that presented infants with two test events: a possible event that accorded with the physical knowledge being examined in the experiment, and an impossible event that violated this knowledge. Given appropriate controls, longer looking at the impossible event was taken to mean that infants (a) possessed the knowledge under investigation, (b) brought to bear this knowledge to form an expectation about the event's outcome, and (c) were surprised when this expectation was violated and sought an explanation for this violation (Baillargeon, 1994).

Over the history of the VOE paradigm, investigators have at times objected to the use of the term "surprise", often on the grounds that infants rarely show prototypical facial expressions of surprise at violations (responses such as facial sobering, bodily stilling, and behavioral freezing are more common; Camras et al., 2002; Scherer et al., 2004). Nowadays, the term surprise is broadly understood in the infancy research community to signal the detection of an expectation violation; we follow this convention here and use the terms surprise and expectation violation interchangeably.

2.3. Knowledge-Consistent and Knowledge-Inconsistent Events

The decade that followed (roughly 1995-2004) brought about two important developments in research on infant cognition; these are described in turn.

2.3.1. Psychological Reasoning

As reliance on VOE and other looking-time tasks (Baillargeon et al., 1985; Leslie & Keeble, 1987), as opposed to action tasks, had revealed unsuspected competencies in young infants' understanding of the physical world, researchers began to ask whether the same might be true for young infants' understanding of the psychological world (i.e., their ability to reason about the actions of agents). Until then, investigations of early psychological reasoning had relied primarily on action tasks (e.g., gaze-following, point-following, social-referencing, or imitative-learning tasks); although many positive results had been obtained with infants in the second year of life (Carpenter et al., 1998), negative results with younger infants were, here again, difficult to interpret and called for alternative tasks that did not require actions.

In the first VOE task to explore early psychological reasoning, 12-month-olds were habituated to a computer-animated event in which two faceless agents, a small and a large circle, stood on either side of a tall barrier until the small circle jumped over the barrier and joined the large circle (Gergely et al., 1995). Following habituation, the barrier was removed, and the small circle either moved to the large circle in a straight line (new-path event) or jumped as before on its way to the large circle (old-path event). Infants looked significantly longer at the old-path than at the new-path event, suggesting that they (a) identified the circles as agents, (b) attributed to the small circle the goal of joining the large circle, (c) expected the small circle to pursue its goal efficiently, and (d) were surprised when the small circle chose an inefficient path to the large circle. This effect was eliminated when the barrier stood out of the path of the small circle in the habituation event; infants could no longer make sense of the small circle's decision to jump, and they refrained from forming further expectations about its actions. Similar results were also obtained with 9-month-olds (Csibra et al., 1999). Although the paradigm used in these reports was referred to as the habituation paradigm, it was evident that expectation violation, rather than novelty, drove test responses: Infants looked longer at the old-path event not because it appeared novel relative to the habituation event—the small circle actually performed the same jumping action in both of these events—but because it was inconsistent with their expectation about how a rational, efficient agent would act in the changed circumstances of the test events.

In another task published at about the same time, it was less immediately evident whether novelty or expectation violation drove infants' responses (Woodward, 1998). In this task, 5- to 9-month-olds were habituated to an event in which an agent faced two different objects, object-A and object-B; across trials, the agent consistently reached for object-A, grasped it, and paused. In the test events, the toys' positions were swapped, and the agent grasped either object-A (old-object event) or object-B (new-object event). Infants looked significantly longer at the new-object than at the old-object event. A novelty-based interpretation of this and similar results (Woodward, 1999) was that infants attended primarily to the object grasped and dishabituated when it changed from familiar object-A to novel object-B. In contrast, an interpretation based on expectation violation was that infants (a) attributed to the agent a preference for object-A over object-B in the habituation trials, (b) expected the agent to continue acting on this preference in the test trials, and (c) were surprised in the new-object event when this expectation was violated. Additional results supported the latter interpretation; for example, infants looked equally at the new- and old-object events if, in the habituation event, object-B was either absent or hidden from the agent (Luo & Baillargeon, 2005a, 2007). These negative results made clear that infants were not merely encoding which object was grasped (had that been true, they would have responded similarly in all cases because the agent always performed the same grasping actions on object-A). It was only when infants could interpret the agent's actions as revealing a preference for object-A over object-B that they looked longer at the new-object event.

To accommodate the findings from these and other investigations of infants' psychological reasoning, the definition of the VOE paradigm, which until then had been narrowly focused on physically possible and impossible events, had to be revised. A broader, more encompassing characterization soon emerged that defined the VOE paradigm as a means of testing *any* knowledge infants might bring to an experiment, be it physical, psychological, or other in nature (Baillargeon, 2004). VOE experiments were now said to present infants with two test events, one consistent with the knowledge being examined in the experiment and one inconsistent with that knowledge. The usual rationale applied: If infants possessed the knowledge under investigation, they should look longer at the knowledge-inconsistent than at the knowledge-consistent event.

2.3.2. Developmental Differences

Initial VOE reports on infants' physical and psychological reasoning were generally positive and brought to light, as we have seen, rich capacities in each domain. However, this same decade (roughly 1995-2004) began to reveal many limitations in these capacities.

In studies of physical reasoning, in particular, at least three broad developmental findings were uncovered (for recent reviews, see Lin et al., 2021, 2022). First, infants often failed to detect subtle, fine-grained violations that could not be discerned without attending to the *featural properties* of objects and their arrangements. The evidence for this finding came from two types of physical violations: *interaction* violations (objects interacted in ways that were not physically possible given their respective properties) and *change* violations (objects spontaneously underwent changes that were not physically possible). For example, although infants as young as 2.5 months were surprised if an object disappeared behind one screen and reappeared from behind another screen, infants under 7.5 months did not detect a violation if an object changed pattern when out of view, and infants under 11.5 months did not detect a violation if an object changed color when out of view (Káldy & Leslie, 2003; Wilcox, 1999; Wilcox et al., 1996).

Second, infants who finally succeeded at reasoning about a particular featural property in one event category often failed to detect interaction and change violations involving the same property in *other event categories.* For example, although by 3.5 months infants were surprised when tall objects became fully hidden behind short screens, similar height violations were not detected until about 7.5 months with short containers, 12 months with short covers, and 14 months with short tubes (Baillargeon & DeVos, 1991; Hespos & Baillargeon, 2001; Wang et al., 2005). Likewise, although 4.5−5-month-olds were surprised if an object changed shape when hidden behind a screen, they failed to detect a violation if the object changed shape when buried in sand (Newcombe et al., 1999; Wilcox, 1999; Wilcox & Baillargeon, 1998b).

Third, infants who finally succeeded at detecting interaction and change violations involving a particular featural property and event category often failed to detect *individuation violations* involving the same feature and category (fewer objects were revealed at the end of an event than were presented during the event, as though one of the objects had magically disappeared). Xu, Carey, and their colleagues were the first to report this baffling failure (Van de Walle et al., 2000; Xu & Carey, 1996; Xu et al., 2004). For example, after two objects that differed only in their featural properties (e.g., two balls that differed in size, pattern, and color) were brought out in alternation from behind a screen, 12-month-olds failed to detect a violation if the screen was lowered to reveal only one of the objects (Xu et al., 2004). Even though by this age most infants can detect interaction and change violations involving size, pattern, and color in occlusion events (Hespos & Baillargeon, 2001; Wang et al., 2004; Wilcox, 1999), infants still failed to detect individuation violations involving these features.

In terms of the VOE paradigm, these findings had several key implications. First, they showed that the paradigm provides a valuable and sensitive tool for examining the protracted and piecemeal development of infants' ability to detect violations. Second, as researchers began to study the mechanisms responsible for this development, it soon became clear that multiple factors contribute to infants' gradual success at detecting violations. With respect to fine-grained interaction and change violations, for example, what matters is not only whether infants possess the relevant *knowledge* but also whether they include the featural information critical for detecting the violations in their *representations* of the events. In the first months of life, event representations tend to be sparse and lacking in featural detail. They become richer over time as infants acquire (through observation and action) rules that identify, for each event category, the features that are causally relevant for predicting outcomes; once identified as relevant, features are routinely included in representations of events from the category. This all means that in a VOE task, infants may fail to detect a fine-grained interaction or change violation not because they lack the relevant knowledge (e.g., objects persist in time and space with all of their physical properties), but because they have not yet learned to include the featural information critical for detecting the violation in their representation of the event (e.g., infants who have not yet learned to include height information when representing containment events cannot be surprised if a tall object becomes fully hidden inside a short container, or if an object surreptitiously changes height when briefly lowered into a container).

Unlike infants' failures with fine-grained interaction and change violations, their failures with individuation violations proved much harder to understand (Leslie et al., 1998; Wilcox et al., 2003; Xu, 2007). A recent account places the blame for these failures on immature *interactions between the two cognitive systems* that work together to track objects (the object-file and physical-reasoning systems), thus extending even farther the range of factors that affect infants’ responses to violations (Lin et al., 2021, 2022; Stavans et al., 2019).

In sum, the steady accumulation of developmental findings led to more nuanced descriptions of infants’ knowledge and the conditions under which it can be observed.

2.4. Expected and Unexpected Events

The following years (roughly 2005-2017) were marked by three main changes. First, VOE tasks were used to assess a broader range of expectations, including sociomoral, biological, numerical, statistical, probabilistic, and linguistic expectations (Buresh & Woodward, 2007; Dewar & Xu, 2007; Hamlin et al., 2007; McCrink & Wynn, 2004; Setoh et al., 2013; Téglás et al., 2007; Thomsen et al., 2011; Xu & Garcia, 2008). Second, just as had happened for physical expectations, researchers using VOE tasks to assess psychological, sociomoral, and other expectations began to vary the number and type of pretest trials administered. Instead of being fully habituated to an introduction or a modulation event, infants might see the event for only a small, fixed number of trials (Xu & Garcia, 2008), or they might receive no pretest trials at all (Sloane et al., 2012). Beyond its practical advantage (fewer pretest trials mean that infants are less fatigued for the test trials), this reduction also reflected the growing acknowledgement that (a) the VOE paradigm is distinct from the habituation paradigm and does not depend on novelty (relative to a familiarized stimulus) to drive infants’ responses, and (b) infants often need only a few repetitions to adequately represent the main subcomponents of events.

The third change was more momentous for the VOE paradigm itself. Until then, the paradigm had been used primarily to assess infants’ knowledge of *veridical rules*. From an adult perspective, veridical rules accurately depict reality; examples are that objects cannot pass through other objects, agents choose efficient paths to their goals, and individuals prefer those who have helped vs. harmed them (Baillargeon et al., 1985; Gergely et al., 1995; Hamlin et al., 2007). This picture began to change, however, with evidence that in the course of learning a complex rule, infants may acquire a series of *incomplete*, *faulty rules* that progressively approximate the more mature rule (Baillargeon, 2002; Baillargeon et al., 2009). In VOE physical-reasoning tasks, this evidence predicted that when shown two physically *possible* events, one consistent and one inconsistent with a faulty rule, infants should be surprised by the latter event and hence should look longer at it than at the consistent event. This made clear, for the first time, that infants could err in VOE tasks by committing not only errors of *omission*, which involved failing to detect a violation, but also errors of *commission*, which involved detecting a violation where there was none (Luo & Baillargeon, 2005b; Wang et al., 2016).

Initial reports of errors of commission built on findings related to young infants' rules about occlusion events. By about 2.5 months, infants acquire a first rule that establishes occlusion as an event category: An object is hidden from view when behind an occluder. At this stage, infants do not yet attend to the shapes or sizes of occluders and expect any object to be hidden when behind any occluder (Aguiar & Baillargeon, 1999). At about 3 months, infants begin to revise this rule: They now expect an object to be visible when behind an occluder whose lower edge is not continuous with the surface on which it rests, creating a low opening between the occluder and the surface (Aguiar & Baillargeon, 2002). At about 3.5 months, infants again revise their rule: They now also expect an object to be visible when behind an occluder that is shorter or narrower than the object (Baillargeon & DeVos, 1991). In line with these findings, 2.5-month-olds committed an error of commission when watching an object move back and forth behind a screen with a low window: They detected a violation when the object appeared in the window (Luo & Baillargeon, 2005b). Similarly, 3-month-olds made an error of commission when watching a tall object move back and forth behind an equally tall screen with a high window: They were surprised when the object appeared in the window (Luo & Baillargeon, 2005b). In each case, infants' faulty rule led them to expect that the object would be hidden when behind the screen.

Additional reports of errors of commission built on findings related to infants' rules about support events. By about 6.5 months, infants expect an object to be stable when released with half or more of its bottom surface on a support (Baillargeon et al., 1992). Over time, infants revise this rule in at least two ways. At about 8 months, they come to realize that an object can be stable with less support as long as the middle of the object's bottom surface is supported (Huettel & Needham, 2000). At about 13 months, infants come to understand that an object that is released with one end on a support may fall even with half of its bottom surface supported, if the object is asymmetrical (e.g., an L-shaped box) and over half of the entire object is off the support (Baillargeon & DeJong, 2017). In line with these findings, 7.5-month-olds were surprised when a rectangular box remained stable with only the middle third of its bottom surface balanced on a narrow support (Wang et al., 2016). Similarly, 11- to 12-month-olds detected a violation when an asymmetrical box fell with half of its bottom surface (but its smaller end) resting on a support (Baillargeon & DeJong, 2017). In each case, infants were surprised by a physically possible outcome that happened to violate their faulty rule.

These findings helped usher the characterization of the VOE paradigm that prevails today (Lin et al., 2022; see also Ginnobili & Olmos, 2021). According to this characterization, VOE tasks are used to assess how infants expect events to unfold, whether these expectations are valid or not. In a typical task, infants are shown two test events, an *unexpected* event that violates a particular expectation and an *expected* event that accords with it. If infants possess this expectation, they will find the unexpected event surprising and will continue processing it, in an attempt to find an explanation for it.

3. Quest for an Explanation

What evidence is there that infants' surprise at an unexpected event typically triggers a quest for an explanation as opposed to, say, a simple state of enhanced attention, interest, or arousal? VOE tasks using total looking time as their measure have yielded several findings that support this assumption (we later discuss related findings with other measures). In this section, we first focus on violations of veridical rules then turn to violations of faulty rules.

3.1. Violations of Veridical Rules

At least three sets of findings support the notion that infants who encounter a violation of a veridical rule in a scene typically search for a way of reconciling what they have observed with their working model of the world. These findings all involve situations in which (a) the observed violation can be explained by *positing an additional, hidden element* in the scene and (b) this explanation does not prompt any significant change to the violated rule itself.

First, as noted earlier, there is evidence that infants no longer show surprise at an unexpected event if they are *given a hint* that an additional, hidden element might be involved in the event. For example, when a tall toy carrot moved back and forth behind a screen with a high window without appearing in the window, 3.5-month-olds did not find this event unexpected if they first received a hint trial in which two tall carrots stood stationary on either side of a windowless screen. At test, when the tall carrot failed to appear in the high window, infants presumably inferred that both tall carrots were used to produce the event, one on either side of the window (Baillargeon & DeVos, 1991). Similarly, when a toy mouse moved back and forth behind two spatially separated screens without appearing in the gap between them, 2.5-month-olds did not view this event as unexpected if the screens were briefly lowered at the start of the trial to reveal a mouse behind each screen (Aguiar & Baillargeon, 1999).

Second, infants show little or no surprise at an unexpected event when they are able to *spontaneously generate an explanation* for it. For example, when a toy mouse moved back and forth behind a screen with a low window without appearing in the window, 3.5-month-olds showed little surprise because they spontaneously inferred that two mice were involved in the event (Aguiar & Baillargeon, 2002; see SOM Figure 1; SOM figures can be found on OSF, 2023; see https://osf.io/k3pae/). The same result was obtained if the screen was lowered at the start of each trial to reveal a mouse and a panel large enough to hide another mouse; when no mouse appeared in the window, infants inferred that a second mouse must have been hidden behind the panel. Infants did show surprise at the event, however, if the screen was lowered at the start of each trial to reveal (a) only a mouse or (b) a mouse and a panel too small to hide a second mouse. Another example comes from an experiment in which 6-month-olds were first introduced to a box that was either inert (inert condition) or self-propelled (self-propelled condition; Luo et al., 2009). In the test event, the box rested behind a screen that was then lifted to hide the box; when raised, the screen occluded the left edge of a second screen. When the first screen was lowered again to reveal no box, infants in the inert condition detected a violation but infants in the self-propelled condition did not, presumably because they inferred that the box had slipped behind the second screen. In line with this interpretation, infants in both conditions viewed the box’s disappearance as unexpected when no second screen was present. Together, these results indicate that in some situations at least, and when the physical layout allows it, infants can spontaneously posit a hidden object or a hidden displacement to make sense of an otherwise unexpected event.[1]

Third, researchers have directly tested infants' ability to generate a plausible explanation for an initial, unexpected event by showing them *a subsequent event that contradicts this explanation*. In one experiment, for example, 4-month-olds were first habituated to an event in which a rod moved back and forth behind two spatially separated screens without appearing in the gap between them (Spelke et al., 1995). In test, the screens were lowered to reveal either one rod (one-rod event) or two rods (two-rod event). Infants looked significantly longer at the one-rod event, suggesting that they made sense of the habituation event by positing the presence of two rods behind the screens. In another experiment, 12-month-olds were habituated to a computer-animated event in which a small agent approached a screen, jumped over the area behind the screen, landed on the other side, and then moved forward until it contacted a larger agent (Csibra et al., 2003). In test, the small agent performed the same actions as before, but the screen was removed at the start of each event to reveal either an obstacle (obstacle event) or empty space (no-obstacle event). Infants looked significantly longer at the no-obstacle event, suggesting that they made sense of the small agent's inefficient jumping action in the habituation event by positing an obstacle behind the screen. In yet another experiment, 7- and 10-month-olds were habituated to an event in which two boxes, each with no back and no top, stood on an apparatus floor; a beanbag was thrown out of one of the boxes and landed between them (Saxe et

---

[1] Readers might be concerned that such results would seem to make difficult the interpretation of negative findings in experimental conditions of VOE physical-reasoning tasks. Did infants show little surprise at the violation being tested because they failed to detect it, or because they could generate a simple explanation for it? In general, there are only a few types of explanations infants are sophisticated enough to consider for violations, so such situations will not often arise. When they do, it should be possible to distinguish between the two interpretations listed above via a modified experimental condition in which infants can no longer posit their explanation (e.g., there is no longer a second screen present). If infants now show surprise at the violation, the second interpretation is more likely; if they still do not, the first interpretation is more likely.

al., 2007). In test, after the beanbag landed, the fronts of the boxes were lowered to reveal a human hand in either the box from which the beanbag had been thrown (same-side event) or the other box (other-side event); in each case, a block occupied the other box. Infants looked significantly longer at the other-side event, suggesting that they made sense of the beanbag's displacement in the habituation event by positing an animate entity in the box from which the beanbag was thrown.

The preceding results paint a consistent picture. When faced with a violation of a veridical rule, infants are sometimes able to make sense of the violation by positing, either spontaneously or via experimental hints, a specific hidden element in the scene (e.g., a duplicate object, a displacement, an obstacle, or an entity capable of exerting force). These explanations do not fundamentally alter infants' rules (e.g., infants would still expect a single object to appear in the gap between two screens, an agent to act efficiently, or an inert object to remain stationary if no force was exerted upon it); however, they may lead infants to elaborate their model of the world to include circumstances that can give rise to apparent violations.

3.2. Violations of Faulty Rules

Evidence that infants who hold a faulty rule and encounter a violation of this rule search for an explanation for this perceived violation comes from experiments inspired by the *explanation-based learning* (EBL) account (Baillargeon & DeJong, 2017; Wang & Baillargeon, 2008). According to this account, when infants are exposed to conflicting outcomes, some consistent and some inconsistent with a faulty rule, they first search for a feature whose values map onto the observed outcomes (e.g., when the feature has value x, the consistent outcomes are observed; when the feature has value y, the inconsistent outcomes are observed). If infants discover such a feature, they then try to build an explanation, using their relevant knowledge, for how it might have contributed to the observed outcomes. If they can build such an explanation, they generalize it, resulting in a revised rule that includes the new feature. If this rule is then verified by a few more empirical exemplars, it is adopted and, from then on, helps guide prediction and prospective action.[2]

Consistent with the EBL account, when 11-month-olds encountered a violation of a faulty support rule in a situation designed to facilitate the EBL process, their search for an explanation resulted in the acquisition of a more advanced rule (Baillargeon & DeJong, 2017). These experiments built on prior findings, described earlier, that when an object is released with one end on a base, 6.5- to 12-month-olds typically expect the object to remain stable as long as the *proportion of the bottom surface* that is resting on the base is greater than that off the base (Baillargeon et al., 1992; Wang et al., 2016). This *proportion-of-contact* rule correctly predicts outcomes for symmetrical objects, but not for asymmetrical objects. Baillargeon and DeJong attempted to induce 11-month-olds to revise this rule in favor of a more sophisticated *proportional-distribution* rule, which is typically acquired at about 13 months: When released with one end on a base, an object remains stable as long as the *proportion of the entire object* that is resting on the base is greater than that off the base.

Infants first watched three pairs of teaching events in which a gloved hand placed the right

---

[2] To be clear, surprise at outcomes inconsistent with faulty rules is only one of the triggers for EBL. Another important trigger is exposure to two or more events that are consistent with infants' current model, have similar representations, and yet yield different outcomes, suggesting that critical featural information is missing from the events' representations and must be added to better predict their different outcomes in the future. For example, infants may observe that when a cover is lowered over an object, sometimes the object becomes fully hidden and sometimes it remains partly visible beneath the cover; although infants do not view either outcome as surprising, the successful identification, via EBL, of a feature that helps explain these different outcomes (e.g., the relative heights of the cover and object) results in the addition of a new rule to their world model (e.g., an object becomes fully hidden under a cover if it is less tall than the cover; Wang & Baillargeon, 2008).

half of an asymmetrical box's bottom surface on a base and then released the box (see SOM Figure 2). Consistent with physical laws, the box fell when released with its smaller end on the base (small-on event), but it remained stable when released in the reverse orientation, with its larger end on the base (large-on event). Each teaching pair involved a different asymmetrical box (a box shaped like a letter B on its back, a right-triangle box, and a staircase-shaped box). Following the teaching trials, infants saw two static test displays in which half of an L-shaped box's bottom surface lay on a base. In the unexpected display, the box's smaller end was supported; in the expected display, the box's larger end was supported. In the teaching trials, infants looked significantly longer at the small-on than at the large-on events overall, thereby committing an error of commission: They perceived the small-on events as unexpected, because these events violated their flawed proportion-of-contact rule (in each small-on event, the box fell even though half of its bottom surface was supported). In the test trials, infants looked significantly longer at the unexpected than at the expected display, suggesting that they did not merely experience surprise when watching the small-on events but actively searched for an explanation for these events. Because the teaching trials were designed to facilitate EBL, infants were able to quickly find such an explanation: They realized that their proportion-of-contact rule was faulty and revised it in favor of a more advanced proportional-distribution rule, which successfully explained why the box fell in the small-on events and exposed the violation in the unexpected display.[3]

---

[3] How did the teaching trials support the EBL process? First, because in each teaching pair the small-on and large-on events differed only in the box's orientation, infants could rapidly zero in on a feature that mapped neatly onto the events' contrastive outcomes: The box fell when the *proportion of the entire box* resting on the base was smaller than that off the base, and it remained stable otherwise. Second, infants could use their physical knowledge to generate an explanation for such a feature: It was plausible that in each teaching trial the base could block the fall of the asymmetrical box when over half of the entire box was on the base, but not when over

This interpretation was supported by control experiments in which the teaching trials were modified to disrupt the EBL process (Baillargeon & DeJong, 2017). For example, if the outcomes of the teaching trials were *reversed* so that each box now fell in the large-on event but remained stable in the small-on event, infants looked equally at the unexpected and expected test displays (see SOM Figure 3). Although infants could see in each teaching pair that the box fell in one orientation but not the other, they were unable to generate a plausible explanation for why this might be the case, and they accordingly retained their proportion-of-contact rule. Infants also failed to acquire the proportional-distribution rule when the boxes used in each teaching pair varied not only in orientation but also in color and pattern, making it difficult for infants to zero in on the feature relevant to the events' differential outcomes (see SOM Figure 4).

The preceding findings provide evidence that the VOE paradigm takes advantage of infants' natural tendency to make sense of the world around them and to look for explanations for unexpected events. In daily life, infants may not often see contrived violations of veridical rules such as those shown in infant laboratories. However, they must often see violations of faulty rules. The findings reviewed above indicate that these perceived violations trigger a quest for an explanation that, under favorable circumstances, can result in improved rules. As Leslie (2004) put it, "a violation of expectation happens when you detect that the world does not conform to your representation of it. Bringing representation and world back into kilter requires representation change, and computing the right change is a fair definition of learning" (p. 418).

Together, the findings we have reviewed on violations of veridical and faulty rules

---

half of the entire box was off the base: The larger, unsupported portion of the box then caused it to tip off the base and topple to the apparatus floor. Armed with this explanation, infants could hypothesize a proportional-distribution rule: An object released with one end on a base will remain stable as long as the proportion of the entire object resting on the base is greater than that off the base. Third, infants could confirm this hypothesized rule because across the teaching trials three different asymmetrical boxes all behaved in accordance with the rule.

highlight some of the key processes in infants' responses to events. When attending to an event in a scene, infants begin by representing the event. They then apply relevant aspects of their world model to this representation, to form an expectation about the event's outcome. If this expectation is violated, they search for an explanation. In the case of veridical rules, successful explanations often involve positing an additional element in the scene, such as a hidden entity or displacement. In the case of faulty rules, successful explanations often involve adding missing featural information to the representation, such as information about the properties or arrangements of the objects in the event. In either case, the added information, interpreted through infants' working model, helps explain the event's outcome.

3.3. Deviations from Impossible Regularities

Imagine that infants are first familiarized with a physically impossible event and then tested with another, physically possible event that deviates from it (following Newcombe et al., 2005, we refer to such conditions as *anomalous* conditions). How would we expect infants to react to this deviation? Our discussion of how infants seek explanations for violations bears on this question. Studies using anomalous conditions indicate that when infants cannot generate explanations for impossible events in familiarization trials, they tend to show little reaction to deviations in test trials. This could be because infants discard their observations of the familiarization events, leaving them with no basis for detecting the deviations in the test events, or because their inability to make sense of the familiarization events causes them to refrain from forming expectations in the test events.

In one study, 5-month-olds watched events in which an object was buried in a wide sandbox (Newcombe et al., 2005). In an anomalous condition, infants saw the same impossible event on five familiarization trials: The object was first buried in location-A (e.g., the midline of the sandbox) and then retrieved from location-B (e.g., 30 cm to the right of the midline; see SOM Figure 5). In the test trial, the object was again buried in location-A and then retrieved from either location-B (consistent event) or location-A (inconsistent event). Infants looked equally at the two events, suggesting that (a) they detected a violation in the familiarization event, as they realized that the object could not spontaneously move through the sand from location-A to location-B, and (b) they showed no particular reaction when next confronted with a deviation from this event. (In an *everyday* condition, the object was buried and retrieved from the same location in the familiarization trials; in line with other evidence of early object permanence, infants looked significantly longer in the test trial if the object was buried and retrieved from different locations as opposed to the same location; see SOM Figure 6).

In another study, 6.5-month-olds watched events in which a cover was lowered over a tall object; the cover was either much shorter or slightly taller than the object (Wang, 2019). In one anomalous condition (see SOM Figure 7), infants saw two impossible events on alternate trials for a total of six familiarization trials: In the short-cover event, the tall object became fully hidden under the cover, and in the tall-cover event, it became only partly hidden under the cover. In the test trials, infants saw the same short-cover event (consistent event) and a new, physically possible tall-cover event in which the object now became fully hidden under the cover (inconsistent event). Infants looked equally at the two events, and Wang offered the following interpretation, inspired by the EBL process described earlier (Baillargeon & DeJong, 2017; Wang & Baillargeon, 2008). At 6.5 months, most infants have not yet identified height information as relevant to covering events, so they typically do not encode the relative heights of the cover and object when watching such events (Wang et al., 2005). However, infants repeatedly watched covering events with two contrastive outcomes, causing them to look for and find a feature whose values mapped onto these outcomes: The object became fully hidden when taller than the cover, but only partly hidden when shorter than the cover. Once this height information was added to their event representations, infants realized that both familiarization events violated their physical knowledge. Being unable to make sense of these events, infants showed no particular reaction when confronted with the inconsistent event. (In a *hint* condition, infants first received a hint trial that helped explain the regularities in the familiarization trials: The tall object could "shrink" or "grow" at will, and it typically shrank when the short cover was used and grew when the tall cover was used. Infants then learned these regularities and looked significantly longer at the inconsistent than at the consistent event; see SOM Figure 8).

Together, the results reviewed in this and the previous sections provide rich evidence that the detection of a violation often triggers a quest for an explanation. Depending on the specifics of the situation, this quest can have weighty consequences for what infants learn as well as for what they fail to learn.

4. Eye-Tracking Technology and the Description of Infants' Looking Behavior

Until now, we have focused on VOE tasks that used total looking time as their main measure. For many years, descriptions of infants' looking behavior in VOE tasks were limited to that measure. Fortunately, for VOE tasks using pre-recorded events such as videotaped or computer-animated events, the advent of infant-friendly eye-trackers made possible (or at least easier) highly detailed descriptions of infants' looking behavior within and across trials (Aslin, 2007; Gredebäck et al., 2010). Eye-trackers can measure not only how long infants look in a trial but also whether they look at specific areas of interest (AOIs), at what point they look at them, how often they look at them, and how long they look at them. Such data can be used in several different ways to glean information about how infants are reasoning about the events they are shown, as we illustrate below.

4.1. Encoding

As events unfold, eye-trackers can provide detailed information about what subcomponents of events infants are attending to; by the same token, eye-trackers can confirm that infants have looked at or encoded subcomponents that are critical for the accurate processing of test events. As an example, consider a computer-animated preference experiment (adapted from Woodward, 1998) that Daum et al. (2012) conducted with 9-month-olds using an eye-tracker. At the start of each of eight familiarization trials, object-A and object-B rested in the top corners of the monitor and a circular screen stood at the center of the monitor, with a small non-human agent (a fish) below it. The agent first moved upward behind the screen; after about 1 s, it reappeared on the far side of the screen, approached object-A, and paused against it. Next, in a static swap trial, the objects were shown in swapped positions, without the agent present. Finally, in the test trials, the agent again moved behind the screen and then approached either object-A (expected event) or object-B (unexpected event) in its new location. Confirming prior reports, infants looked significantly longer at the unexpected than at the expected event. Moreover, detailed analyses focusing on the AOI corresponding to each object indicated that during the familiarization and test trials, infants looked longer at whichever object the agent approached. During the swap trial, infants looked equally at the two objects, suggesting that they had the opportunity to notice their swapped positions.

4.2. Anticipation

In VOE tasks where the expected and unexpected outcomes involve different locations, eye-trackers can reveal whether infants correctly anticipate the expected outcome (i.e., look to its location before it occurs). So far, studies that have measured anticipatory responses have produced somewhat mixed results. In the task of Daum et al. (2012) described above, for example, most infants correctly anticipated that the agent would approach object-A in the last two familiarization trials, but few infants did so in the first test trial. Other preference tasks, however, have elicited better anticipatory performances (Cannon &Woodward, 2012; Kim & Song, 2015; see also Southgate & Begus, 2013).

In one task, 11-month-olds received four trials that each had three phases involving different movie clips (Cannon & Woodward, 2012). In the first phase, object-A and object-B rested in the right top and bottom corners of the monitor; an agent’s hand entered from the left, moved straight across the scene, and deflected just past midline to grasp object-A; this event was repeated three times. In the second phase, the object’s positions were swapped. Finally, in the third phase, the hand moved as before but paused just past midline. At that point, infants were more likely to make their first look from the hand to object-A, and this pattern was evident from the first trial onward. Infants thus attributed to the agent a preference for object-A in the first phase, and they anticipated that she would reach for object-A in its new location in the third phase. Kim and Song (2015) obtained converging results with 6-month-olds, again using videotaped events. In six familiarization trials, an agent wearing a visor sat centered behind object-A and object-B; in each trial, she consistently reached for object-A. Next, in the swap trial, the agent was absent, and the objects were shown in swapped positions. Finally, in the test trial, the agent simply sat behind the swapped objects for a 6-s period. An offline frame-by-frame analysis of this period (sometimes referred to as a “poor man’s eye-tracker”) showed that infants anticipated that the agent would reach for object-A in its new location.

Together, these studies suggest that infants sometimes do and sometimes do not actively anticipate an expected outcome by looking at the right AOI at the right time. As recent controversies surrounding the anticipatory-looking method have made clear, whether infants show robust anticipation in a task appears to depend on subtle factors that are far from perfectly understood (Baillargeon et al., 2018; Grosse Wiesmann et al., 2018; Kulke et al., 2018; Schuwerk et al., 2022). As a case in point, although the studies described above all used a preference task, they differed in multiple respects, making it difficult to pinpoint which differences mattered for eliciting correct anticipation.

4.3. Inference

Eye-tracking data can provide evidence that when watching an event, infants build a representation of it online and integrate new information as it becomes available. In a computer-animated experiment conducted by Cesana-Arlotti et al. (2018), 12- and 19-month-olds received test trials in which two different animated objects (e.g., a snake and a ball, each with an identical rounded red top) entered a scene and paused to the left of an animated cup. After a screen hid the two objects, the cup went behind the screen, scooped up one of the objects, and then returned to its initial position; although the top of the object protruded from the cup, infants could not determine which object it was as both objects had identical top parts. Next, one object (e.g., the snake) emerged to the right of the screen, paused briefly, and then returned behind the screen. Finally, an object again emerged to the right of the screen and paused into view; this was either the same object as before (e.g., the snake; expected event) or the other object (e.g., the ball; unexpected event). Across test pairs, infants looked significantly longer at the unexpected than at the expected event, suggesting that (a) when the first object emerged from behind the screen, they inferred, by exclusion, which object was in the cup (e.g., “so it must be the ball that is in the cup!”), and (b) when that same object next emerged from behind the screen, they were accordingly surprised. Support for this interpretation came from additional analyses that examined the proportion of test trials in which infants shifted their gaze, when the first object emerged into view, from this object to the cup. At both 12 and 19 months, the proportion of test trials with object-to-cup shifts significantly predicted infants' overall surprise at the unexpected events. These and control data suggested that as each test event unfolded, infants integrated new information, as it became available via perception or inference, into their representation of the event, and this updated representation guided their responses to the event's final outcome.

5. Extensions of the VOE Paradigm

Over the past decade or so, researchers' reliance on infants' looking behavior in VOE tasks has been greatly enriched by the addition of a whole host of measures targeting other responses. In this section, we first describe five of these measures and then reflect on how they have helped broaden our conceptual understanding of how infants respond to violations.

5.1. Pupil Dilation

Pupil dilation is an automatic physiological response caused not only by changes in luminance but also, to a lesser degree, by cognitive factors such as the detection of unexpected stimuli (Laeng et al., 2012). As eye-trackers automatically measure pupil size many times per second, infancy researchers have begun to use pupil dilation as an index of surprise in VOE tasks. One advantage of this measure is its high temporal resolution: It can provide detailed information about the time-course of infants' response to an unfolding unexpected event and, in particular, pinpoint how soon after the violation infants give evidence of having detected it (Hepach & Westerman, 2016; Zhang & Emberson, 2020).

In one study, 6-month-olds saw two events in which an agent spooned food on the back of a recipient's hand, instead of in her open mouth (Gredebäck & Melinder, 2010). In the expected event, a barrier prevented the agent's access to the recipient's mouth; in the unexpected event, the barrier was again present but no longer blocked access. Infants showed greater pupil dilation (over a 2.4-s period that covered the movement of the spoon and contact with the recipient's hand) in the unexpected than the expected event. Likewise, in another study, 12-month-olds saw two events in which an agent sat behind a screen, which was then lowered to reveal an empty location (Pätzold & Liszkowski, 2019). In the expected event, the agent simply sat quietly before the screen was lowered; in the unexpected event, the agent pointed excitedly at the area behind the screen, as though enthralled by something there. Infants showed greater pupil dilation (over a 4-s period after the screen was lowered) in the unexpected as opposed to the expected event. In both studies, infants thus showed greater pupil dilation when an agent acted irrationally as opposed to rationally.

5.2. Event-Related Potentials

Electrophysiological responses assessed through encephalogram (EEG) recordings can also index surprise in VOE tasks and provide precise information about the time course of infants' response to an unfolding event. Building on prior evidence that error detection in adults is associated with brain activity with negative polarity over middle-frontal areas (Gehring et al., 1995), Berger et al. (2006) presented 7-month-olds with videotaped addition and subtraction events adapted from Wynn (1992). In some trials, for example, a screen was lifted to hide an object, an experimenter's hand added another object behind the screen, and the screen was finally lowered to reveal either one object (unexpected outcome) or two objects (expected outcome). The authors analyzed the brain activity of the infants who looked longer at the unexpected than at the expected outcomes overall, focusing on event-related potentials (ERPs) over frontal areas between 330 and 530 ms after the lowering of the screen in each trial. Significantly greater negativity was found for the unexpected as opposed to the expected outcomes, leading the authors to conclude that “the association of looking time with frontal activity related to error detection suggests that […] looking time at the very least indicates that infants have detected a violation of their expectations” (p. 12651).

Other reports have focused on one particular ERP component, the N400: In adults, detection of a semantically incongruous stimulus is associated with brain activity with negative polarity over parietal areas approximately 400 ms after the onset of the stimulus (Kutas & Federmeier, 2011). Building on these results, Parise and Csibra (2012) presented 9-month-olds with a computer monitor displaying an upright screen; their mother, seated at their side, stated which object was behind the screen (e.g., “a duck!”), and then the screen was lowered to reveal either the labeled object (expected outcome) or some other object (unexpected outcome). Analyses of infants’ ERPs over parietal areas after the object was revealed indicated that infants exhibited larger N400s when shown the unexpected as opposed to the expected outcomes. Infants thus expected to see the object labeled by their mother and detected a violation when they did not. In related experiments testing early false-belief understanding (Forgács et al., 2020), 14-month-olds also produced larger N400s when an agent who held a false belief about the identity of a hidden object (e.g., she believed it was an apple, but it had been replaced with a toy car) uttered a label that matched the current object (e.g., “a car!”). Infants thus detected a violation when the agent’s label (though accurate from their perspective) was incongruous with her belief about the object’s identity.

5.3. Event Selection

Another measure of expectation violation in VOE tasks involves giving infants a forced choice between watching again either an unexpected or an expected event; the rationale is that if infants detect the violation in the unexpected event, they should be more likely to choose to see that event again, to process it further as they attempt to make sense of it. In one report using an event-selection (or infant-triggered-video) task, 8-month-olds faced two computer monitors and were first shown that touching each monitor triggered a different event (Jin et al., 2018). One monitor displayed an expected event: A woman folded laundry on one side of a room until a baby lying in a stroller at the back of the room began to cry; the woman then walked to the stroller and rocked it gently, as though trying to comfort the baby. The other monitor displayed an unexpected event that was similar, with two exceptions: A different woman was involved, and when the baby began to cry, she walked to a chair located next to the stroller to pick up more laundry to fold, thus ignoring the crying baby. Next, each monitor displayed a still picture from its event, and infants chose which event they wanted to see again by touching the corresponding monitor; while the selected event played, additional touches to either monitor had no effect. Across trials, infants were more likely to select the unexpected event, suggesting that they detected a violation when the woman ignored the crying baby and were trying to find an explanation for this violation.

5.4. Exploration of Violation Objects

Another measure of expectation violation that can be used in VOE tasks involving physical violations examines whether infants are more likely to approach and explore an object that was previously involved in a violation than an object that was not. With appropriate controls, greater exploration of the violation object is taken to indicate that infants detected the violation they were shown and, through their exploratory activities, are trying to glean information that can help them make sense of this violation (Sim & Xu, 2017; Stahl & Feigenson, 2015). For example, Stahl and Feigenson (2015) showed 11-month-olds a single live event in which a target object behaved in a manner that either accorded with or violated the solidity principle (the object stopped against an obstacle or passed through it) or the gravity principle (the object remained stable with or without support). Next, infants were presented with the target object and a novel distractor object to explore for 60 s. Infants who had seen an unexpected event spent more time exploring the target object than infants who had seen an expected event, providing evidence that they were surprised by the object's behavior and were searching for an explanation. In line with prior findings, this effect was eliminated if, following the unexpected event, infants were given a hint that helped explain how the event had been produced (e.g., the obstacle was rotated to reveal an opening at its center; Perez & Feigenson, 2022).

Additional results indicated that when searching for explanations, infants not only spent more time exploring violation objects but also tailored their exploratory activities to the specific violations shown (Stahl & Feigenson, 2015). Thus, infants who had seen an object pass through an obstacle were more likely to bang it (as though testing its solidity), whereas infants who had seen an object remain suspended in midair were more likely to drop it (as though testing its sensitivity to gravity). Similar results were obtained in a task tapping a faulty physical rule as opposed to a veridical one (Zhang & Wang, 2019). When a rectangular target box remained stable with its middle third balanced on a narrow base, 7.5-month-olds (whose faulty proportion-of-contact rule stated that an object could not remain stable with less than half of its bottom surface supported) spent more time exploring this target box than another box. Infants also spent more time specifically lifting and dropping the target box, as though searching for a way to make sense of what they had observed.

Finally, yet other results (Stahl & Feigenson, 2015) indicated that infants were also quicker to learn an incidental fact about a violation object (e.g., it squeaked when shaken) than about a non-violation object, again pointing to infants' enhanced interest in learning about the violation object and collecting information that might help explain its behavior.

5.5. Downstream Responses to Violation Agents

Other VOE measures focus on downstream consequences of infants' explanations for violations. In the case of psychological and sociomoral violations, infants often seem to take these violations to reveal internal deficiencies, either in rationality (in the case of psychological violations) or in moral character (in the case of sociomoral violations). As these explanations carry unfavorable evaluations that color infants' subsequent expectations, attitudes, and behaviors toward the violation agents, VOE measures targeting these downstream consequences focus on negative responses such as avoidance, distrust, vigilance, and indirect punishment.

5.5.1. Psychological Violations

In the psychological-reasoning literature, many studies of infants' downstream responses to violation agents have focused on epistemically unreliable agents, who act in a manner inconsistent with information that is or should be available to them. A typical task involves two successive contexts: The first introduces an unreliable agent, and the second examines infants' responses to the agent in a different situation. For example, after seeing an unreliable agent express facial and vocal excitement ("wow!") when looking inside empty containers, 14-month-olds were less likely to peek around a barrier when the agent expressed excitement as she looked behind the barrier (Chow et al., 2008), and 16-month-olds did not find it unexpected if the agent searched in the wrong location for an object she had watched being hidden (Poulin-Dubois & Chow, 2009). Similarly, after seeing an unreliable agent use everyday objects in an atypical manner (e.g., putting sunglasses on his foot), 14-month-olds were less likely to learn a novel conventional action demonstrated by the agent that involved activating a light-box by touching it with the forehead (Zmyj et al., 2010). Finally, after observing an unreliable agent label familiar objects incorrectly (e.g., label a ball a "rabbit"), 16-month-olds were less likely to request information from the agent about novel objects (Begus & Southgate 2012), and 18-month-olds were less likely to learn a novel label taught by the agent for a novel object (Brooker & Poulin-Dubois, 2013). In sum, after observing an agent's epistemic violation, infants did not find it unexpected if the agent gave further evidence of epistemic unreliability, and they tended to view the agent as less trustworthy or credible: They were less likely to seek new information from the agent or to learn new actions or labels taught by the agent.

5.5.2. Sociomoral Violations

In the sociomoral-reasoning literature, several studies on downstream responses to violation agents have focused on wrongdoers who fail to adhere to moral principles such as fairness or ingroup support. For example, after watching two distributors divide resources, one fairly and one unfairly, 17-month-olds preferred the fair over the unfair distributor (Geraci & Surian, 2011; for a review and meta-analysis, see Margoni & Surian, 2018); 10-month-olds expected an informed bystander to give a treat to the fair as opposed to the unfair distributor (Meristo & Surian, 2013); 15-month-olds looked longer at the unfair than at the fair distributor upon hearing disembodied praise (e.g., "she's a good girl!"), as though they realized that these utterances applied only to the fair distributor (DesChamps et al., 2016); and 25-month-olds were more likely to return a dropped ball the distributors were playing with to the fair as opposed to the unfair distributor (Surian & Franchin, 2017). Similarly, after watching a wrongdoer repeatedly harm an ingroup victim (e.g., destroy a tower, puzzle, and drawing made by the victim), 25-month-olds did not find it unexpected if the wrongdoer next divided resources unfairly between two ingroup or outgroup recipients; however, they did find it unexpected if the wrongdoer next acted generously by giving an ingroup member most of a resource to be shared between them (Ting & Baillargeon, 2021; see also Surian et al., 2018). In sum, after observing an agent's moral violation, infants and toddlers did not find it unexpected if the agent gave further evidence of a flawed character, they were less likely to prefer or help the agent, and they expected others to share their negative evaluation and to refrain from giving treats or praise to the agent.[4]

5.6. Comparing Measures

The varied array of measures used in VOE tasks highlights three successive phases in infants' responses to violations, whether real or perceived. When an event does not unfold as expected, infants detect a violation (*detection*). They then continue processing the event (more than they would for a comparable but expected event), to try to make sense of what has happened (*processing*). Finally, in some cases at least, infants succeed in generating an explanation for the violation they have observed (*explanation*).

Sorting the different VOE measures we have discussed in terms of these phases, we can see that pupil dilation and ERPs are concerned mainly with the detection phase and provide time-course information about how soon infants give evidence of detecting a violation after receiving the necessary information to do so. Total looking time, event selection, and object exploration are concerned mainly with the processing of detected violations. Depending on the situation, this processing takes different forms: If an unexpected event repeats continuously, infants watch multiple loops as they try to make sense of it; if an unexpected event ends with a paused scene,

---

[4] Some physical-reasoning tasks can also be construed as focusing on the downstream consequences of infants' explanations. As an example, consider a support task (Hespos & Baillargeon, 2008) in which 5.5- and 6.5-month-olds were shown two identical toys at the center of a wall, one supported from below and one not. At both ages, infants reached preferentially for the supported toy: They inferred that the unsupported toy must be attached to the wall (explanation), and hence they viewed it as irretrievable and eschewed it in favor of the other, retrievable toy (downstream consequence).

infants continue looking at the scene as they process what has happened; if given the opportunity to watch again either the same unexpected event or another, expected event, infants choose the unexpected event; and when given the opportunity to approach and explore either a violation object or another, non-violation object, infants choose the violation object. Whatever its form, this processing serves to help infants find an explanation for the violation they have detected. Finally, measures of downstream consequences are concerned mainly with infants' explanations for violations. In the social domain, explanations for agents' violations often carry unfavorable evaluations and, as such, have negative consequences for infants' expectations, attitudes, and behaviors toward the agents.

6. Postdictive, Integrative, and Predictive Processing

The advances described in sections 4 and 5 have helped us better understand how infants process events and respond to outcomes that violate their expectations. These same advances have also helped shed light on the issue of *when infants typically generate their expectations for events' outcomes*, and we turn to this issue below.

6.1. Three Possibilities

To make our discussion more concrete, imagine that infants see four familiarization events followed by a single test event. Assuming that infants possess the knowledge necessary to correctly reason about this event, when might they form their expectation about its outcome? One can envision at least three possibilities.

A first possibility is that as infants watch the test event unfold, they simply store a shallow representation of it in memory. After the event ends, they look back on it mentally: They interpret it using their world model and relevant information from the familiarization events, and they form an expectation about what the event's outcome should have been. Mismatches between this expectation and what actually happened elicit surprise. Following Wellman (2011), we refer to this possibility as *postdictive* processing; infants “proceed backward” (p. 37) after the event ends, so the key steps of interpretation and expectation formation do not occur until then.

Another possibility is that as soon as the test event begins, infants get to work interpreting it, again using their world model and information from previous trials. Infants may also form partial predictions about *recurring* aspects of the event, based on these trials. Crucially, however, they are unable to form predictions about *new* aspects of the event (e.g., due to insufficient time, limited information-processing resources, and so on). Thus, infants in a rotating-screen task may be able to predict that the screen will rotate again—but not that it will stop when it reaches the box that now stands in its path. Similarly, infants in a preference task may be able to predict that the agent will again reach for one of the toys before her—but not that she will reach for her preferred toy in its new location. In this second possibility, much more processing occurs while the event is unfolding; nevertheless, infants still form an expectation about the event’s outcome only after it ends. We refer to this possibility as *integrative* processing; as the event unfolds, much information is integrated into its interpretation, but the critical step of expectation formation still occurs only after the event ends.

Yet another possibility is that infants engage in *predictive* processing: In addition to interpreting the test event as it unfolds, they form an expectation about its final outcome *before* it takes place (for a similar distinction between integrative and predictive processing in speech perception, see Federmeier, 2007). To return to our same examples, this third possibility means that infants predict that the screen will stop rotating when it reaches the box, or that the agent will reach for her preferred toy in its new location, before they observe these outcomes.

6.2. Evaluation of the Three Possibilities

Which of the three possibilities just described best capture(s) when infants form expectations about events' outcomes? The findings reviewed in sections 4 and 5 bear on this question in several ways. First, the findings cast doubt on the notion that infants' processing of events is primarily postdictive. We have seen that under favorable circumstances, infants update their representation of an unfolding event as soon as information becomes available that licenses particular inferences (Cesana-Arlotti et al., 2018); they anticipate an event's outcome even the first time they see the event (Cannon & Woodward, 2012; Kim & Song, 2015); and they give evidence of having detected an unexpected outcome (e.g., via pupil dilations or ERPs) very soon after they receive the information necessary to detect it (Berger et al., 2006; Forgács et al., 2020; Gredebäck & Melinder, 2010; Parise & Csibra, 2012). These findings call into question the possibility that infants' reasoning about events in VOE tasks is entirely postdictive. To the contrary, events appear to be processed as they unfold, and this makes adaptive sense. From a computational standpoint, processing an event *incrementally* and adding to its representation as new pieces of information become available (via perception or inference) helps distribute the work of interpretation across the entire event. In contrast, processing the event only after it ends adds to its information-processing costs and runs the risk of overwhelming infants' limited cognitive resources.

Second, infants' failure to anticipate an event's outcome does not necessarily mean that they did not predict or expect that outcome. Whereas behavioral anticipation implies prediction, the reverse is not true: Prediction can occur without anticipation. For example, an infant in a preference task might fully expect the agent to approach her preferred toy at the start of the unexpected event but still not focus on that AOI prior to the agent's action; instead, the infant might look at the unpreferred toy because that is the location the agent last approached, or the infant might simply look at the agent and wait for her to act (in the same way, an adult might be able to predict a partner's actions in a situation but not bother to anticipate them). As noted previously, infants may actively anticipate an expected outcome—by looking at the right AOI at the right time—only when features of the situation strongly elicit that response, and researchers are just beginning to understand what some of these features may be (Téglás & Bonatti, 2016; Schuwerk et al., 2022).

Third, the evidence that infants sometimes anticipate outcomes in VOE tasks makes clear that predictive processing is already present in infancy (Cannon & Woodward, 2012; Gredebäck et al., 2018; Kim & Song, 2015; Téglás & Bonatti, 2016). Such processing has many advantages. First, as is widely acknowledged, predictive processing plays a key role in supporting the interpretation of incoming signals in all aspects of cognition (Clark, 2013; Farmer et al., 2013; Köster et al., 2020; Zhang & Emberson, 2020). As Clark (2013) put it, "brains are essentially prediction machines" that are "constantly attempting to match incoming sensory inputs with top-down expectations or predictions" (p. 181). Second, predictive processing is essential not only for interpreting the world around us but also for responding to it in a timely and prospective manner. For infants, being able to predict how objects will move or interact can result in smoother and more skillful actions (von Hofsten, 2007). Similarly, being able to predict how social partners will act, or what the outcomes of their actions will be, can result in more seamless and finely tuned interactions (Brandone et al., 2014). Third, prediction errors contribute to learning. When learners use their current model of the world to generate precise expectations about what will happen next, they receive targeted feedback on the accuracy of their model; error signals can prompt revisions and updates that can lead to better predictions about the world, and more competent interactions with it, in the future (Köster et al., 2019, 2020).

Together, the preceding points highlight the need to keep conceptually separate surprise responses, predictions, and anticipations in discussions of VOE tasks. A surprise response to an unexpected outcome can occur only post-hoc, after infants have experienced and detected the violation in the outcome; but the fact that this response must occur post-hoc does not mean that infants' *processing of the event* must also occur post-hoc. Infants may form a prediction about an event's outcome and later show, post-hoc, a surprise response when this prediction is violated. Moreover, infants who predict an event's outcome may or may not bother to actively anticipate it. Predictions do not always translate into anticipations, which require mental effort and may be worth doing only when they yield some cognitive advantage (Téglás & Bonatti, 2016).

6.3. Looking Forward

Our discussion suggests that infants who successfully detect a violation in an unexpected event may do so in two different ways. One involves *integrative* processing: For one reason or another, infants are unable to predict the event's final outcome, though they may form partial predictions about other aspects of the event. After they experience the event's outcome, they assess whether it accords with their world model, and they respond with surprise if it does not. The other way involves *predictive* processing: As the event unfolds, infants predict its final outcome, using relevant aspects of their world model, and they respond with surprise when what happens next does not match their prediction. It seems plausible that depending on the specifics of the situation and their own attentional state, infants may fluidly move from one form of processing to the other, even from trial to trial within a testing session.

The distinction between integrative and predictive processing raises many interesting directions for future research. First, new measures (beyond visual anticipation) are needed to ascertain, on a trial-by-trial basis, whether infants who are watching an event are engaging in predictive or only integrative processing. Second, research is needed to identify under what conditions each form of processing is more likely. Finally, research is needed to explore the cognitive consequences of each form of processing. For example, are infants more likely to detect the violation in a complex unexpected event if they are able to predict its outcome? Does an inability to form a predictive expectation sometimes impair infants' ability to detect a violation, because processing costs become too high?

7. Strengths of the VOE Paradigm

Our detailed discussion of how the VOE paradigm has evolved and broadened over time makes it easier to appreciate its strengths.

First, the rationale of the VOE paradigm is intuitive and takes advantage of infants' natural tendency to make sense of the world around them. In a typical task, infants are shown two test events, an expected event that accords with a particular expectation and an unexpected event that violates it. With appropriate controls, surprise at the unexpected event is taken to mean that infants possess the expectation under investigation, detect the violation in the event, and attempt to make sense of this violation so as to better predict outcomes in the future.

Second, the VOE paradigm is highly versatile in terms of the nature of the expectations it can assess. It can be used to study expectations from many different content areas (e.g., physical, psychological, sociomoral, biological, numerical, statistical, probabilistic, and linguistic expectations); it can be used to assess veridical or faulty expectations; and it can be used to track changes in expectations with age or with experimental manipulations designed to teach, prime, or otherwise induce specific expectations.

Third, the VOE paradigm is remarkably flexible in several respects. It can be used with children aged 2 months to 3 years, despite the marked differences in their perceptual, cognitive, motor, and linguistic abilities. It can be used with a wide range of measures, as we have seen (e.g., total looking time, pupil dilation, ERPs, event selection, exploration of violation objects, and downstream responses to violation agents). It can be used with varying numbers and types of pretest trials (e.g., introduction, modulation, and hint trials) and with varying numbers of test trials, in either between-subjects or within-subjects designs. It can be used with live events as well as with videotaped or computer-animated events. Finally, in the social domain, it can be used with non-verbal events as well as with events that include age-appropriate verbal information, and it can be used with either human agents or artificial non-human agents such as animated objects, puppets, and two-dimensional characters (for a discussion of the use of artificial agents in infant studies, see Kominsky et al., 2022).

Fourth, in many areas of cognition, there is substantial converging evidence between VOE tasks using different measures as well as between VOE and non-VOE tasks. In the social realm, for example, evidence that infants in the second year of life can attribute false beliefs to others comes not only from VOE tasks using ERPs (Forgács et al., 2020), as we have seen, but also from VOE tasks using total looking time (Onishi & Baillargeon, 2005; Scott et al., 2015) and from non-VOE tasks (Buttelmann et al., 2009; Scott et al., 2012). Similarly, in the physical domain, evidence that by the second year infants can individuate objects from different basic-level categories (e.g., a ball and a toy duck) comes from VOE tasks using total looking time (Wilcox & Baillargeon, 1998a; Xu & Carey, 1996) as well as from non-VOE tasks (Cacchione et al., 2013; Van de Walle et al., 2000).

Fifth, the VOE paradigm can be used not only in laboratory settings but also in field sites, making possible comparisons between infants in Western and traditional, non-Western cultures. As an example, a VOE false-belief task using total looking time, first administered to 18-month-olds in an infant laboratory in the Midwestern United States (Scott et al., 2010), was subsequently administered to 16- to 30-month-olds in a Salar community in northwest China and to 17- to 30-month-olds in a Shuar community in southeastern Ecuador (Barrett et al., 2013). Results were similar in all three cultures.

Sixth, although most VOE reports focus on group data, in a few reports individual differences in infants' responses to unexpected as opposed to expected events have been found to predict responses in other tasks, at either the same age or later ages. Thus, in the social realm, 7-month-olds' performance in a preference task (Woodward, 1998) was positively associated with their performance in an imitation task in which an experimenter either grasped one of two objects or touched it with the back of her hand (Hamlin et al., 2008): Infants who looked proportionally longer at the new-object event in the first task were more likely to choose the object grasped (but not touched) by the experimenter in the second task, suggesting that both tasks assessed the same capacity to encode agents' intentions (Thoermer et al., 2013). In another report, performance in a preference task at 7 months was related to performance in a moral intention-understanding task (Killen et al., 2011) at 5 years: Infants who looked proportionally longer at the new-object event were subsequently more likely to evaluate the intention of an accidental transgressor as positive (e.g., Tim thought he was doing something good when he mistook Max's paper bag for trash and threw it away; Sodian et al., 2016). In the physical realm, infants who exhibited the strongest response to a solidity violation (an object passed through an obstacle) at 11 months were found to also exhibit the strongest response to a support violation (an object remained stable without support) at 17 months (Perez & Feigenson, 2021). Moreover, infants' performance at 17 months predicted their explanation-based curiosity at 3 years, as assessed by a parental survey (e.g., "my child is bothered when he/she does not understand something, and tries to make sense of it"); no such link was found for novelty-based curiosity (e.g., "my child is attracted to new things in his/her environment").

Seventh, the VOE paradigm can be used not only with human infants but also with non-human animals (henceforth animals). VOE tasks have been employed with a wide range of animals including birds, cats, chimpanzees, dogs, dolphins, elephants, lemurs, lions, meerkats, monkeys, and sea lions, some tested in the wild and others in the laboratory (Drayton & Santos, 2018; Gilfillan et al., 2016; Singer & Henderson, 2015; Takagi et al., 2016; Völter et al., 2023; for reviews, see Ginnobili & Olmos, 2021; Winters et al., 2015). Total looking time is often used to measure animals' surprise, but pupil dilation, exploration, and vigilance are also used. The expectations examined include physical, numerical, probabilistic, psychological, and sociomoral expectations. Results are sometimes similar to those obtained with human infants, and sometimes not. To illustrate, Bird and Emery (2010) found that rooks (birds of the crow family) could detect three types of support violations that human infants aged 6 months and older have been shown to detect (Baillargeon, 1995). Their study took advantage of rooks' natural tendency to look through small holes: When peering through a hole in a box, perched birds could see static unexpected or expected support displays presented on a monitor at the back of the box. Across experiments, birds looked significantly longer when an object remained stable (a) in midair, with no contact with a base, (b) against the side of a base, with no support from below, or (c) with only a small portion of its bottom surface resting on top of a base. In contrast, Cacchione and Krist (2004) found that chimpanzees who saw videotaped events depicting these same support violations looked significantly longer at the (a) and (c) violations, but not the (b) violation, and Murai et al. (2011) confirmed these results with chimpanzees and Japanese monkeys. These results illustrate some of the fundamental questions raised by comparative VOE tasks: Are the expectations uncovered in human infants and in animals similar or different in nature? Are these expectations formed and revised through similar or distinct processes? And what evolutionary, structural, and environmental factors contribute to observed differences?

Eight, research on human adults' reactions to physical violations has also revealed intriguing parallels with infant VOE findings (Lewry et al., 2021; Liu et al., 2023). For example, in online surveys, Lewry et al. (2021) showed adults drawings of violations previously used with infants and asked them to rate how interesting or surprising the violations were. Results revealed two complementary trends. First, adults rated violations that are typically detected at very young ages (because they do not require the representation of detailed featural information) as more interesting than violations that are typically detected at later ages (after infants have learned to include the necessary featural information in their event representations). Thus, a violation in which an object appeared to pass through a barrier was deemed more interesting than a violation in which a tall object became hidden inside a short container or under a short cover (even though all three violations involved the same principle of persistence; Lin et al., 2022). Second, just as infants' surprise at violations dissipates when they can generate an explanation for them, adults rated violations that were easily explained as less interesting than violations that were more difficult to fathom. Thus, a violation in which an object remained stable against the side of a base (presumably via adhesives or magnets) was deemed much less surprising than one in which an object remained stable in midair, with no contact with a base (for further discussion of adult reactions to violations used with infants, see Baillargeon et al., 2018, and Low & Edwards, 2018).

Finally, in the field of computer science, artificial-intelligence (AI) systems are being trained to acquire common-sense, human-like physical rules through exposure to relevant physically possible events. To assess AI systems' learning performance, VOE tasks similar to those used with human infants are sometimes used as an evaluation tool: Following training on specific rules, AI systems are asked to distinguish between (untrained) possible and impossible events that bear on the same rules (Piloto et al., 2022; Smith et al., 2019). In addition, VOE benchmarks are being developed to ascertain whether AI systems are acquiring meaningful physical rules that can transfer to new events or more superficial expectations (Dasgupta et al., 2023; Riochet et al., 2018; Weihs et al., 2022). Results so far have been mixed, suggesting that much more needs to be done to create AI systems capable of human-like physical reasoning.

8. Limitations of the VOE Paradigm

Alongside its many strengths, the VOE paradigm possesses several limitations. This is, of course, not particular to the VOE paradigm but is true, in one way or another, of all paradigms used to investigate early cognitive development. In this section, we discuss four broad limitations of the VOE paradigm, along with some of the strategies that have been used to overcome them.

8.1. Interpreting Positive Findings

By itself, evidence that infants in an experimental condition of a VOE task respond differentially to the unexpected and expected events is never sufficient to draw firm conclusions. This is because such evidence leaves unclear the basis of infants' response. As we pointed out in the Introduction, positive VOE findings may be open to conceptual as well as non-conceptual interpretations, and conceptual interpretations themselves may vary from richer to leaner or shallower interpretations. To draw the conclusion that infants are responding on the basis of a rich conceptual expectation, researchers must show that infants produce a significantly different response in one or more additional conditions designed to elicit such responses; for ease of communication, we refer to these as control conditions. Broadly speaking, control conditions follow one of two approaches. One is to test alternative, low-level interpretations of infants' responses to the test events, as in *no-expectation* or *shallow-expectation* controls. The other is to show that conceptually relevant changes to the test events modulate infants' responses in a plausible and coherent way, as in *conceptual-variation* controls.

8.1.1. No-Expectation Controls

No-expectation controls are designed to address the possibility that infants in an experimental condition may look longer at the unexpected event not because it violates their world model (conceptual interpretation), but because it presents a low-level feature that attracts and holds their attention (non-conceptual interpretation). For example, relative to the expected event, the unexpected event might involve a more eye-catching arrangement, motion, pattern, color, or other featural property. To evaluate such a non-conceptual interpretation, a no-expectation control typically (a) preserves the low-level feature hypothesized to drive infants' response in the experimental condition but (b) introduces a modification that, under the conceptual interpretation, should no longer present infants with an expectation violation. The rationale is that if the non-conceptual interpretation is correct, infants should show the same looking pattern as in the experimental condition, because the low-level feature is still present; if the conceptual interpretation is correct, however, infants should now look equally at the two test events, because neither event violates their world model. (A no-expectation control that does not preserve the low-level feature or does so imperfectly will, of course, be less useful at ruling out the non-conceptual interpretation).

To illustrate, in an experiment testing early sensitivity to fairness, 9-month-olds saw live test events in which a human distributor brought in and divided two items between two identical animated penguin puppets (Buyukozer Dawkins et al., 2019, adapted from Sloane et al., 2012).

In front of each puppet was a placemat; the distributor placed one item on one puppet's placemat, and then she placed the second item either on the same puppet's placemat (2:0 event) or on the other puppet's placemat (1:1 event). Infants looked significantly longer if shown the 2:0 vs. the 1:1 event. A conceptual interpretation of this finding was that infants expected the distributor to divide the two items fairly and hence detected a violation in the 2:0 event. In contrast, a non-conceptual interpretation was that infants simply found the final asymmetrical display in the 2:0 event more eye-catching than the symmetrical display in the 1:1 event. To evaluate this possibility, additional infants were tested in a no-expectation control identical to the experimental condition except that the puppets were inanimate (i.e., they no longer moved or talked). Infants now looked equally at the 2:0 and 1:1 events. Similar results were obtained by Meristo et al. (2016) in a computer-animated experiment with 10-month-olds: Infants looked significantly longer at the 2:0 than at the 1:1 event when the recipients were animated triangles with faces, but they looked equally at the events when the recipients were inanimate baby bottles. In each experiment, results in the no-expectation control differed significantly from those in the experimental condition, casting doubt on the notion that a low-level preference for asymmetry drove infants' response.

8.1.2. Shallow-Expectation Controls

Shallow-expectation controls are designed to address the possibility that infants in an experimental condition may look longer at the unexpected event not because they possess the rich conceptual expectation under investigation (rich interpretation), but because a leaner or shallower expectation elicits the same response (lean interpretation). Shallow-expectation controls are often similar to no-expectation controls: A critical change is introduced to the experimental condition; under the lean interpretation, infants' response should be the same as before, because the shallow expectation still applies; under the rich interpretation, infants should now look equally at the two events, because neither violates their model of the world.

As an example, consider once again the fairness experiment of Buyukozer Dawkins et al. (2019). An alternative, shallower interpretation of the experimental condition was that infants looked longer at the 2:0 event not because they expected the distributor to divide the items fairly between the two potential recipients, but because they expected similar individuals to have similar numbers of objects. To evaluate this possibility, additional infants were tested in two conditions (adapted from Sloane et al., 2012). The *cover-experimental* condition was identical to the experimental condition except that the distributor removed covers resting over the puppets' empty placemats, one at a time, before dividing the two items either unequally or equally between them. The *cover-control* condition was similar except that the distributor no longer brought in and divided the two items: In each event, she simply removed the covers to reveal the items already resting on the puppets' placemats. Infants in the cover-experimental condition again looked significantly longer at the 2:0 than at the 1:1 event, but infants in the control-cover condition looked equally at the two events, casting doubt on the notion that infants simply expected similar individuals to have similar numbers of objects. Meristo et al. (2016) obtained similar results: When the distributor simply revealed as opposed to distributed the items, by pushing off-screen a small barrier resting below each animated triangle, infants looked equally at the 2:0 and 1:1 outcomes. Together, these results indicated that infants expect windfall resources to be divided fairly between similar individuals, but hold no particular expectation about how many resources similar individuals may already have in their possession.

8.1.3. Conceptual-Variation Controls

Another approach to supporting a rich conceptual interpretation of an experimental condition is to show that conceptual changes to the condition modulate infants' expectations in a predictable manner. Finding significantly different yet coherent responses across conceptual variations can help bolster not only the rich interpretation offered for the experimental condition but also the broader theoretical framework from which this interpretation is drawn.

To illustrate, consider experiments that built on the fairness experiments described above to examine whether 19-month-olds would expect ingroup support to prevail over fairness when resources were limited (Bian et al., 2018). In test trials adapted from Buyukozer Dawkins et al. (2019), the human distributor was replaced with another animal puppet, and the group memberships of the distributor and recipient puppets were manipulated. When the distributor (e.g., a monkey) divided two items between two recipients of a different group (e.g., two giraffes), infants expected a fair distribution, as before. However, when one of the recipients belonged to the same group as the distributor (e.g., one monkey distributor, one monkey recipient, and one giraffe recipient), infants looked significantly longer when the outgroup recipient received both items (favors-outgroup event) or when each recipient received one item (equal event) than when the ingroup recipient received both items (favors-ingroup event). Because there were only enough items for the distributor's group (e.g., two items and two monkeys), infants expected ingroup support to prevail, and they detected a violation when the outgroup recipient received either one or both items. In line with this interpretation, when the distributor brought in three items, gave away two, and left with the third one, infants now looked significantly longer at the favors-ingroup or favors-outgroup event than at the equal event. As there were enough items to go around, fairness was expected to prevail, and infants detected a violation if either recipient was given both items.

Although for the sake of clarity we have discussed no-expectation, shallow-expectation, and conceptual-variation controls separately, they often overlap in what they accomplish. In the fairness studies described above, for example, the shallow-expectation control could also serve as a no-expectation control, and the conceptual-variation control could also serve as a no-expectation and a shallow-expectation control. More generally, these various controls help illustrate that the careful evaluation of alternatives to rich conceptual interpretations not only is essential to support such interpretations but can result in a fuller and more nuanced understanding of infants' expectations.

8.2. Interpreting Negative Findings

Infants' failure to show surprise at an unexpected event often reveals limited competencies: Infants may not yet possess the necessary knowledge, or they may not yet represent the necessary featural information. However, there are also situations where infants do possess the competencies needed to detect the violation in an unexpected event and yet fail to do so. Below, we list an array of factors that can contribute to such failures. Understanding these factors is important not only to gain a better grasp of the VOE paradigm but also to highlight sources of variation that can negatively impact attempted replications, leading to inconsistent results.

8.2.1. Events

Infants may fail at a VOE task because the events shown are somehow inappropriate. One difficulty can be that the events do not take into account limitations in infants' *perceptual abilities*. In one study, 4.5-month-old girls were surprised when an object rolled through an obstacle behind a screen, but 4.5-month-old boys were not (Baillargeon & DeVos, 1991): Because stereoscopic depth perception matures more slowly in boys, they could not determine whether the obstacle stood in or out of the path of the object when the screen was lifted at the start of the event. When shown a spatial layout that circumvented these difficulties, however, both 2.5-month-old girls and boys detected the violation (Spelke et al., 1992). Similarly, 9-month-olds succeeded at the computer-animated detour task of Csibra et al. (1999), but 6-month-olds did not: Because their ability to infer depth information from two-dimensional images was still immature, they had difficulty interpreting the animations. When shown animations that provided richer depth cues, however, they readily succeeded (Csibra, 2008).

Even when infants correctly perceive events, other difficulties can still arise. One issue may be *content*: Events that imperfectly target the expectation under investigation, because they contain ambiguous or superfluous elements, will confuse infants and leave them uncertain about what is happening. Another issue may be *pacing*: Infants may fail at a task because events unfold too rapidly for them to take in what they are seeing. Yet another issue may be *processing load*: Infants may fail to detect a violation in an unexpected event because this event requires them to process too much novel information at once (i.e., infants are given “too much to chew”). In such a case, adding pretest trials to introduce subcomponents of the test events ahead of time may improve infants’ performance. All of these issues make clear that producing a good VOE task is in some ways akin to producing a good movie, play, or magic show: The audience must be able to follow events well enough to bring to bear relevant aspects of their world model and form expectations about what will happen next.

To illustrate how event-related issues can affect infants’ performance, consider VOE tasks that were designed to assess 15-month-olds’ understanding of pretense (Onishi et al., 2007). In one experiment, infants received a test trial in which an experimenter turned two cups right-side up (to show they were empty), lifted a visibly empty jug, and pretended to pour into one of the cups. Next, the actor pretended to drink from either the “filled” cup (expected event) or the other cup (unexpected event). Infants looked significantly longer if shown the unexpected as opposed to the expected event, suggesting that they were able to evaluate the consistency of the experimenter's pretend actions. However, this effect was eliminated if the cups were replaced with less prototypical containers such as shoes. The authors conjectured that this atypical substitution might have given infants too much information to process at once. Supporting this hypothesis, infants succeeded in detecting the violation in the unexpected event if they first received a familiarization trial in which the experimenter pretended to drink from a shoe ("she drinks from shoes!").

8.2.2. Task Structure

Infants may fail at a VOE task because the number of pretest or test trials administered creates difficulties. We have just seen that receiving *too few pretest trials* may make it hard for infants to process all of the novel information they encounter in the test trials. The opposite is also true: Receiving *too many pretest trials* can cause infants to shift to a shallower form of processing ("yah, yah, same old thing"); if the test trials are similar to the pretest trials, this superficial, perseverative processing can lead infants to overlook critical changes and fail at the task. In such a case, reducing the number of pretest trials, or eliminating them altogether, may improve infants' performance. Another issue has to do with receiving *too many test trials*. When an unexpected event is complex or challenging, infants may be able to form an expectation about its outcome the first time they see it, but not thereafter. In such a case, focusing on the data from the first test trial or using a single-test design may give a truer sense of infants' competence.

In addition to the numbers of pretest and test trials infants receive, the *consistency between the formats of these trials* can also matter. If the pretest trials all involve repeating event loops, but the test trials each present a single event loop that ends with a paused scene, infants may fail simply because they are puzzled as to why the events suddenly "stop" or "freeze", and this distracts them from processing the contents of the events.

Another issue related to task structure has to do with the *need to use a between-subjects design* for presenting the test events. In some VOE tasks, infants may be able to handle two or more test trials, but only if a between-subjects design is used, with the same event being shown across trials. At times, this may be due to working-memory limitations and confusability (e.g., if one test event involves a container and another involves an otherwise identical tube, infants who see both events in alternation may forget as each event progresses whether they are looking at the container or the tube). At other times, a between-subjects design may be required due to contamination between two test events, when seeing the first event renders the second one physically impossible (e.g., a display is revealed to be composed of two adjacent objects in the first event, but to be composed of a single object with two distinct parts in the second event). In such cases, the use of a between-subjects design, with one or more test trials, can provide a better assessment of infants' capabilities.

A final issue that combines some of the preceding ones is showing infants *too many different test trials*. In many VOE tasks, infants are able to handle seeing two different test events (e.g., an expected and an unexpected event) in alternation for one or more pairs of trials; seeing more than two different test events, however, can cause difficulties. The introduction across trials of multiple differences among events can confuse infants and make it harder for them to zero in on critical changes. Using a between-subjects design to show separate groups of infants one or two of the events typically yields more interpretable data.

To illustrate how task structure can affect infants' performance, consider experiments that found perseverative responding in a VOE task (Aguiar & Baillargeon, 2003; see SOM Figure 9). In one experiment, 6.5-month-olds received four familiarization trials in which an experimenter’s gloved hand first held a ball (attached to the bottom of a rod) above a short container that was wider than the ball; a screen briefly hid these objects and then was removed to reveal the ball resting inside the container. Next, infants received three pairs of test trials identical to the familiarization trials except that tall containers were used, one wider than the ball (expected event) and one only half as wide as the ball (unexpected event); in either case, the screen was removed to reveal the ball’s rod protruding from the container. Despite the fact that by 6.5 months relative width has already been identified as a causally relevant feature for containment events (Wang et al., 2004), infants looked equally at the two test events. The authors conjectured that infants might have been led by the overall similarity of the familiarization and test trials to shift to a shallower mode of analysis that caused them to perseveratively apply the same expectation as before (“the ball will fit into the container”) to each test event. Supporting this hypothesis, infants succeeded in detecting the violation in the unexpected event if the familiarization trials were modified either by removing the container from the familiarization trials (displacement condition) or by removing the back half of the container so that it now served as an occluder (occlusion condition). Either way, infants no longer saw containment events in the familiarization trials, and this led them to categorize the test events as new events that called for new expectations.

8.2.3. Testing Sessions

Infants may fail at a VOE task because they have *already received the task* in a prior testing session conducted days, weeks, or even months previously. This may cause infants’ attention to be divided between (a) processing the events before them and forming an expectation about their outcomes and (b) remembering when and where they have seen these events before (e.g., like adults watching a TV show who are preoccupied with figuring out whether they have seen that particular episode before). Another issue is that infants may fail at a VOE task that is preceded, in the same testing session, with one or more other tasks involving the same agents. *Contamination across tasks* may lead to negative results, as we saw previously: Recall, for example, that after seeing an unreliable agent express excitement when looking inside empty containers, infants did not find it unexpected if she next searched in the wrong location for an object she had watched being hidden (Poulin-Dubois & Chow, 2009).

Yet another issue is that giving infants *multiple tasks* in the same testing session, even when there is little danger of contamination across tasks, can sometimes lead to spurious negative results in a VOE task, due to simple mental fatigue. Infants tire much faster than do other children and adults, in part because most cognitive tasks require more mental effort from them. Obtaining positive results when a VOE task is administered either alone or first, but negative results when it is administered amidst a suite of other tasks, tells us that infants can demonstrate a particular expectation under the most optimal conditions, but not otherwise.

8.2.4. Infants

As with most tasks, infants who are quiet, alert, and attentive during a VOE task are more likely to succeed than infants who are not. VOE reports typically list exclusions of infants due to *state-related reasons*, such as being overly fussy, drowsy, active, or distracted (e.g., by a sock). Infants may also be excluded if they are highly inattentive or disengaged from the task.

Even when infants appear quiet and alert, another issue that can affect their performance in a VOE task is how relaxed they feel in the novel environment of the lab. It seems plausible that infants who feel more stressed, due to *temperamental tendencies or attachment-related experiences*, would perform less well. Although this issue has received little experimental attention to date, there is some evidence that infants whose mothers provide a secure base for exploration are better able to process and interpret what they observe in the lab than are infants with a less secure attachment. In one experiment, Hohenberger et al. (2012) showed 10-month-olds computer-animated expected and unexpected collision events adapted from Kotovsky and Baillargeon (1994). Infants with more sensitive mothers looked significantly longer at the unexpected than at the expected event, as in prior findings. In contrast, infants with less sensitive mothers looked equally at the events, most likely because they were too stressed to fully process the novel animations they were shown and detect the violation in the unexpected event.

8.3. Interpreting Reversed Findings

As we have seen throughout this article, unexpected events typically elicit stronger responses—whether assessed by total looking time, pupil dilation, ERPs, event selection, or object exploration—than do expected events.[5] Nevertheless, VOE tasks sometimes yield *reversed* patterns, with infants showing stronger responses to expected than unexpected events. This unusual pattern most likely reflects additional or alternative processes beyond surprise at an expectation violation. To our knowledge, only a small handful of VOE experiments have

---

[5] Findings by Kidd et al. (2012) are sometimes taken to cast doubt on this conclusion. According to these findings, infants look away sooner from events whose probability is either very high (non-surprising events) or very low (highly surprising events), compared to events of intermediate probability. In one study, 7-month-olds received 42 trials on an eye-tracker. Each trial involved up to 30 reveals: A screen was raised (1 s) and lowered (1 s) continuously to reveal either an object (e.g., a ball) or no object, according to the pre-selected probability for the trial. Trials ended when infants looked away for 1 s. The results obtained suggested that infants who looked away in a trial at the 11th reveal, for example, were more likely to do so (a) if the object had been present on all previous reveals and was again present or (b) if the object had been present on all previous reveals but was now absent, compared to (c) if the object had been present or absent about an equal number of times. Given the specifics of the task (e.g., infants could see up to 1,260 reveals across trials; stimuli appeared and disappeared at the start and end of each trial, undermining beliefs in their permanence; and there was no evidence that infants who looked away from a reveal had actually noticed the presence or absence of the object behind the screen), the task is quite distinct from traditional VOE tasks. It may be better construed as a pattern-detection task in which infants, quite plausibly, look away sooner from less variable as opposed to more variable patterns.

reported reversed findings to date; in this section, we describe three such experiments and offer suggestions about what additional or alternative processes might be involved. We acknowledge from the outset that our suggestions are tentative and exploratory; we hope they will bolster constructive efforts to understand what special circumstances might lead to reversed findings in VOE tasks as well as how researchers might experimentally manipulate these circumstances to gain further insights into how infants represent, reason about, and respond to events.

8.3.1. Showing Vigilance toward a Violation Agent

Vigilance toward a violation agent—or enhanced attention to a potential threat—can sometimes masquerade as longer looking to an expected event. In experiments by Meristo and Surian (2013, 2014), 10-month-olds were familiarized to two computer-animated events: In one, a fair distributor divided two strawberries equally between two recipients, and in the other, an unfair distributor divided the strawberries unequally between the recipients. In the test events, the distributors stood in the bottom corners of the monitor, and a new character entered at the top. In one condition, the character gave a strawberry to the fair (expected event) or the unfair (unexpected event) distributor; in another condition, the character took away a strawberry from the fair (unexpected event) or the unfair (expected event) distributor; in yet another condition, the character hit the fair (unexpected event) or the unfair (expected event) distributor three times. In the first condition, infants looked significantly longer at the unexpected than at the expected event; in the other two conditions, infants showed the reverse pattern, looking longer at the expected event. One possible interpretation of these findings is that because the character was absent during the familiarization trials and hence was uninformed about the distributors' actions, infants spent little time forming expectations about what it would do. Instead, they tended to focus on the *unfair* distributor: In all three conditions, they looked longer, out of vigilance, whenever it was approached, to glean further information about it. Support for this interpretation comes from a recent experiment in which infants saw the same familiarization events followed by novel test events in which each distributor appeared alone, moving in and out of an enclosure (Margoni et al., 2023). Infants looked significantly longer when shown the unfair distributor, reinforcing the notion that vigilance toward that violation agent guided their responses (see also Vaish et al., 2008).

Further results (Meristo & Surian, 2013) made clear that vigilance toward the unfair distributor guided infants' test responses only if the character was *absent* during the familiarization trials, so that its behavior in the test trials could not be interpreted as a reaction to the distributors' actions. If the character was *present* during the familiarization trials, infants considered its potential reactions in the test trials, in ways consistent with other findings in the sociomoral-reasoning literature (Choi & Luo, 2015; Sloane et al., 2012). Specifically, if the character was present *and could see the distributions,* infants looked significantly longer if it gave a strawberry to the unfair (unexpected event) as opposed to the fair (expected event) distributor. In contrast, if the character was present *but could not see the distributions* because a barrier blocked its view, infants looked equally at the two events, as though they recognized that, to the uninformed character, either distributor would seem deserving of the strawberry. More generally, these results indicated that whenever the character was present in the familiarization trials, infants tended to focus in the test trials on making sense of its behavior as opposed to showing vigilance toward the unfair distributor, underscoring the importance for infants of making sense of chains of events as they unfold.

8.3.2. Avoiding a Distressing Unexpected Event

Focus on an expected as opposed to an unexpected event may occur when the latter event is deemed by infants to be not only surprising but also emotionally distressing or aversive, causing them to avoid it. As a case in point, consider recent VOE event-selection experiments in which 12-month-olds were presented with computer-animated events involving two groups of geometric characters with faces, squares and circles (Jin & Baillargeon, 2023). Infants chose between an unexpected event (e.g., a large square holding a stick hit a small square with it three times) and an expected event (e.g., a large circle holding a stick jumped next to a small circle three times). Infants were significantly more likely to choose the unexpected event, suggesting that they were surprised by the ingroup-support violation depicted and wanted to watch it again to try to make sense of it. When the number of hits or jumps in the events was increased from three to six, however, infants' response pattern reversed: Infants were now significantly more likely to select the expected event, suggesting that they viewed the unexpected event as not only surprising but also emotionally distressing or aversive, and they accordingly chose to watch the other, more neutral event instead. Both of these effects were eliminated when the large characters were swapped across events (e.g., the large circle hit the small square, and the large square jumped next to the small circle), so that the events now depicted interactions between members of different groups or moral circles and hence between potential predators and prey (e.g., foxes and chickens).[6]

---

[6] The notion that infants may avoid unexpected events they view as emotionally distressing suggests a possible interpretation of an experiment by Johnson et al. (2010) in which 13-month-olds were shown a computer-animated adult-baby separation scenario. At the start of each test event, a large oval (the "adult") stood half-way up a steep hill, and a small oval (the "baby") cried at the bottom of a hill. Next, the adult either returned to the baby (expected event) or continued to the top of the hill (unexpected event). Securely attached infants looked significantly longer at the unexpected than at the expected event, whereas insecure infants showed the reverse (albeit only marginally significant) looking pattern. One interpretation of this last finding could be that the insecure infants tended to look away from the unexpected event because their prior experiences with their caregivers led them to view it as emotionally distressing (Dykas & Cassidy, 2011). However, another interpretation, suggested by the findings of Hohenberger et al.

8.3.3. Managing Two Distinct Streams of Processing

Focus on an expected event may also occur when something about the situation causes infants to simultaneously pursue two separate streams of processing (or lines of thought), each with its own purpose; under these challenging circumstances, focus on the expected event may simply make it easier to manage both streams at once. As a case in point, consider an experiment by Kuhlmeier et al. (2003) in which 12-month-olds saw computer-animated events involving three faceless characters: a circle, a triangle, and a square. Infants were habituated to two events. In one, the circle attempted to climb a hill without success and was then helped to the top of the hill by one character (e.g., the triangle). The other event was similar except that the circle was pushed to the bottom of the hill by the other character (e.g., the square). In the test events, the hill was removed, the triangle and the square stood in the top corners of the monitor, and the circle initially stood centered at the bottom of the monitor. Next, the circle moved to the middle of the monitor, wavered briefly (as though trying to decide where to go), and finally approached either the triangle (expected event) or the square (unexpected event). Infants looked significantly longer at the expected than at the unexpected event. One possible interpretation of this reversed finding is that infants were attempting to process two distinct questions in the test events: whether the circle was likely to approach the triangle or the square, and whether the triangle and the square were facing front or back and could detect the circle's silent approach. Attempting to glean information about this second question while also forming an expectation about the first question might have led infants to dismiss the unexpected event and focus on the expected event, which imposed a smaller processing load overall (i.e., it made sense, given what had happened in the familiarization events, that the circle would approach the helpful triangle, and it mattered less

---

(2012) reviewed earlier, might be that the insecure infants were simply too stressed in the novel environment of the lab to fully process the test events.

whether the triangle was or was not aware of this approach).

Two sets of results provide support for this interpretation. First, subsequent experiments in which the characters were given eyes elicited the standard looking pattern: Infants now looked significantly longer at the unexpected than at the expected event (Hamlin et al., 2007; Lee et al., 2015). The addition of eyes, of course, rendered the second question moot, as it made clear which direction the triangle and the square were facing. Infants could simply focus on forming an expectation about where the circle would go, and they were surprised when their expectation was violated. Second, in another experiment by Kuhlmeier and colleagues cited in Wynn (2007), infants saw identical test trials preceded by different familiarization trials adapted from Woodward's (1998) preference task. These familiarization trials were identical to the test trials except that the positions of the triangle and square were swapped. After seeing the circle consistently approach the triangle as opposed to the square, infants expected the circle to continue approaching the triangle in the test trials, and they detected a violation when it approached the square instead (for similar results with other faceless agents, see Johnson et al., 2007; Luo & Baillargeon, 2005a). In this task, the triangle and the square were simply inanimate objects (there was no evidence to suggest otherwise), so the question of which way they were facing never arose. Here again, infants could simply focus on forming an expectation about where the circle would go, and they were surprised when their expectation was violated.

8.3.4. Additional Remarks

In this section, we have considered possible explanations for reversed findings from three VOE tasks, all from the social domain. We do not wish to claim that this constitutes an exhaustive list of possible explanations for reversed findings, or that such findings are unlikely to arise in other domains, such as the physical domain. Some years ago, in pilot work, one of the authors (R.B.) found that infants tested with simple support events tended to look longer at the expected than at the unexpected event if the two events were accompanied by a salient melody; this pattern reversed when the melody was removed. Though anecdotal, these findings are consistent with the possibility raised above that reversed findings can occur when the experimental situation inadvertently causes infants to pursue two separate streams of processing—in this case, one focused on the physical possibility of each support event and one on the accompanying melody (e.g., to process it more fully or to judge whether the experimenter's actions were synchronized with the melody).

8.4. Comparing Positive Findings

In general, the VOE paradigm provides only a simple yes/no answer to the question of whether infants possess a particular expectation: As a binary testing method, it does not support the rank-ordering or proportional comparison of positive findings across tasks or conditions. To illustrate, imagine that three groups of infants were tested with a different physical-, psychological-, or sociomoral-reasoning VOE looking-time task, and that each group detected the violation they were shown; the data collected could not be used to rank-order the three violations from most to least surprising. As Aslin and Fiser (2005) put it, "Yes-No paradigms use quantitative data (e.g., looking times) to draw qualitative conclusions" (p. 94). A differential score (e.g., looking time to the unexpected event minus that to the expected event) that was twice as large in task-A as opposed to task-C would not mean that the violation in task-A was twice as surprising or involved an expectation that was twice as entrenched in infants' world model. With limited sample sizes and limited data from individual infants, variations in positive findings cannot support quantitative conclusions that some violations are more surprising than others.[7]

As Aslin and Fiser (2005) pointed out, quantitative conclusions are possible in other paradigms used with infants, such as psychophysical studies of visual acuity: Multiple data points are collected from each infant under three or more conditions, and then these data are used to generate a function linking the stimulus variable (e.g., the width of black-and-white stripes) to the response of interest (likelihood of detecting the stripes). Unfortunately, as we have seen in our discussion of VOE limitations, extensive testing of individual infants in traditional VOE tasks is often unproductive (e.g., repeated trials can cause infants to shift to more superficial or perseverative responding, and repeated sessions can trigger memory searches that interfere with VOE responding). The situation may change in the future if new measures of surprise or new experimental approaches enable researchers to obtain sufficient data from individual infants to make quantitative comparisons across tasks or conditions meaningful (see also Byers-Heinlein et al., 2021; DeBolt et al., 2020).

9. Challenges Leveled at the VOE Paradigm

The VOE paradigm has been the target of criticism since its inception. In this section, we discuss six challenges that have been leveled at the paradigm. In keeping with the rest of our article, we focus on theoretical criticisms rather than on methodological concerns. While such concerns are beyond the scope of our article, we have no doubt that the VOE paradigm, like other infant paradigms, will benefit from the many ongoing efforts to improve the robustness and replicability of research practices in developmental psychology (Byers-Heinlein et al., 2021;

---

[7] The same is true when comparing different conditions within a task. For example, evidence of a larger differential score in a false-belief than a true-belief condition (or the reverse) could not be taken as evidence that infants found it somewhat easier to reason about agents' false as opposed to true beliefs (or the reverse). The difference between the two conditions could be due to random variation, or it could reflect design decisions (e.g., number and contents of pretest trials) that enhanced performance in one condition more than in the other.

Davis-Kean & Ellis, 2019; Eason et al., 2017; Frank et al., 2017; Kominsky, 2022; Margoni, 2022; Margoni & Shepperd, 2020; Rubio-Fernandez, 2019; Schuwerk et al., 2022).

9.1. Novelty and Familiarity Preferences

One criticism of VOE tasks is that what appear to be surprise responses to unexpected events may be no more than transient familiarity or novelty preferences induced by pretest events in the tasks (Bogartz et al., 1997, 2000; Cashon & Cohen, 2000; Thelen & Smith, 1994).

This criticism is derived from research on the factors that modulate whether infants display familiarity or novelty preferences in habituation and other familiar-novel tasks. When presented with visual stimuli, infants often show a familiarity preference under shorter familiarization conditions and a novelty preference under longer familiarization conditions, and the amount of familiarization needed to elicit a novelty preference varies with age and task difficulty (Hunter et al., 1982, 1983; Hunter & Ames, 1988; Rose et al., 1982). To illustrate, Rose et al. (1982) presented infants with an object until they accumulated 5, 10, 15, 20, or 30 s of looking. Next, the object was paired with another object of a different shape. At 3.5 months, infants preferred the familiar object after 10 s of familiarization and the novel object after 30 s; at 6.5 months, they preferred the familiar object after 5 s and the novel object after 15 s. The authors concluded that "as infants begin to process a stimulus, they prefer to look at that which is familiar; once processing is more advanced, their preference shifts to that which is novel" (p. 711). It makes sense that infants whose processing of an object is interrupted would be motivated to return to it to complete their processing; for infants to recognize and categorize objects in everyday life, adequate representations must be stored, and infants who erratically flitted from object to object would have difficulty forming such representations. As Hunter et al. (1982) put it, infants give priority to "consolidating information they are in the process of acquiring before moving on to make new discoveries" (p. 528).

Can VOE findings be mere familiarity or novelty preferences? When we introduced Woodward's (1998) preference task earlier in this article, we noted that its findings were open to a *novelty-based* interpretation: In the habituation trials, infants attended mainly to the object grasped by the agent, object-A; in the test trials, they dishabituated when this object changed from familiar object-A (expected event) to novel object-B (unexpected event). However, recall that additional data refuted this novelty-based interpretation: Infants looked equally at the test events if in the habituation event object-B was either absent or hidden from the agent by a screen (Bíró et al., 2011; Choi et al., 2022; Luo & Baillargeon, 2005a; Luo & Johnson, 2009). Other early VOE tasks were open to *familiarity-based* interpretations. In the rotating-screen task of Baillargeon et al. (1985), for example, infants were habituated to a screen that rotated back and forth through a 180^o arc, in the manner of a drawbridge. Next, a box was placed in the screen's path, and the screen rotated through either the same 180^o arc as before (unexpected event) or a novel, shorter arc that ended against the box (expected event). Similarly, in the detour task of Gergely et al. (1995), infants were habituated to a small agent who jumped over a barrier to join a large agent. Next, the barrier was removed, and the small agent traveled to the large agent using either the same circuitous path as before (unexpected event) or a novel, straight path (expected event). In both tasks, the unexpected event was thus more similar than the expected event to the habituation event, raising the possibility that infants looked longer at the unexpected event simply because they were continuing to process the full rotation or circuitous path shown in the habituation event. Here again, however, control data did not support these familiarity-based interpretations. As mentioned earlier, infants did not look longer at the screen's familiar 180^o rotation if the box was absent or out of the screen's path in the test events (Baillargeon, 1987; Baillargeon et al., 1985), and they did not look longer at the small agent's familiar circuitous path if no barrier blocked its path in the habituation event (Csibra et al., 1999; Gergely et al., 1995).

Because infants in VOE tasks must represent the objects and events they are shown to reason about them, encoding and consolidation processes no doubt contribute to these representations. However, transient familiarity or novelty preferences are unlikely to play a major role in infants' test responses. First, as we have just seen, control data typically do not support such interpretations (see also Csibra et al., 2003; Kamewari et al., 2005; Phillips & Wellman, 2005; Sodian et al., 2004). Second, VOE tasks with only test trials preclude such interpretations (Bian et al., 2018; Lin et al., 2021; Stavans et al., 2019; Wang et al., 2004). Third, some of the findings reviewed in previous sections provide specific evidence against such interpretations. For example, recall that in the rotating-screen task of Baillargeon (1987), 3.5-month-olds who habituated quickly to the screen's 180° rotation looked significantly longer at the unexpected than at the expected event, whereas infants who failed to habituate (presumably because they had difficulty encoding the screen's rotation) looked equally at the two events; this is the opposite looking pattern than that predicted by a familiarity-based interpretation. Recall also that in the burial task of Newcombe et al. (2005), 5-month-olds who were familiarized to a possible burial event (i.e., an object was buried and found at the same location) reacted with increased looking if next tested with an impossible burial event (i.e., the object was buried and found at different locations), but this effect was eliminated in an anomalous condition in which the two events were presented in the reverse order (see SOM Figures 5 and 6). As the authors noted, this asymmetrical looking pattern makes clear that infants did not simply acquire a specific expectation in each condition about where the buried object was likely to be found, as a novelty-based interpretation would have predicted. Instead, infants brought to bear their physical knowledge and responded differently in the two conditions.

Finally, and most generally, all of the evidence reviewed in this article supports the notion that positive responses in VOE tasks (as opposed to familiar-novel tasks) are driven primarily by expectation violations. As we have seen, when watching an event, infants may not only form an expectation about its outcome but also actively anticipate this outcome; when an expectation is violated, they may reveal their surprise via different measures (e.g., pupil dilation, ERPs, total looking time, and event selection); when searching for an explanation of a physically impossible outcome, they may engage in various exploratory activities, including targeted manipulations of violation objects; when faced with a violation of a faulty physical rule, their quest for an explanation, in circumstances favorable to EBL, may result in the acquisition of a more advanced rule; and when their explanation for the behavior of a violation agent carries a negative evaluation of the agent’s rationality or moral character, downstream consequences can include avoidance, distrust, and lowered expectations. Together, these findings provide converging evidence that VOE responses are driven primarily by infants’ propensity for making sense of the world so that it becomes more predictable.

9.2. A Myriad of Alternative Low-Level Accounts

Another criticism of the VOE paradigm is that positive findings are open to a myriad of alternative, low-level accounts that do not grant infants sophisticated expectations and reasoning abilities. As a consequence, researchers who offer rich conceptual interpretations of their results can never be confident in their chosen interpretations. To be clear, the criticism here is not that low-level accounts *can* be offered for VOE findings; rather, the criticism is that for any VOE finding, even if one or two deflationary accounts can be successfully ruled out, *dozens* more remain, undermining confidence in any conclusion drawn from the finding (Aslin & Fiser, 2005; Haith, 1998).

This criticism is in some ways correct, but it is instructive to reflect on what makes it so: Numerous low-level accounts become possible for VOE findings when considerations of basic cognitive plausibility are ignored. To illustrate this point, we discuss two low-level accounts that have been proposed for VOE findings (Bogartz et al., 1997; Heyes, 2014). Both accounts, we argue, make far-fetched assumptions about how infants process events.

9.2.1. Thin-Segment Account

In a VOE task described earlier (Baillargeon & Graber, 1987), 5.5-month-olds saw familiarization events in which a tall (tall-rabbit event) or a short (short-rabbit event) rabbit moved back and forth across a scene, briefly disappearing from view as it passed behind a large screen. The test events were identical except that the screen now had a high window; each rabbit moved back and forth as before without appearing in the window. Infants looked significantly longer at the tall-rabbit than at the short-rabbit test event, and the original interpretation of this finding was that infants expected the tall rabbit (who was taller than the window's lower edge) to appear in the window and were surprised that it did not. An alternative, *thin-segment* account (Bogartz et al., 1997) was that in each familiarization and test event, infants focused on the rabbit's face and processed only the thin horizontal segment of the screen at the same level as the face. In the tall-rabbit test event, this segment happened to include part of the window and hence was novel in appearance relative to the familiarization events. Infants thus looked longer at the tall-rabbit than at the short-rabbit test event simply because they detected the presence of the window in the former but not the latter event.

Additional findings (discussed earlier) obtained with the rabbit task and similar tasks involving a high or a low screen window cast doubt on the validity of the thin-segment account (Aguiar & Baillargeon, 1999, 2002; Baillargeon & DeVos, 1991; Baillargeon & Graber, 1987; Luo & Baillargeon, 2005b). Instead of reiterating these findings, however, here we focus on the *lack of cognitive plausibility* of the account. To see what we mean, consider two assumptions the account had to make to explain the findings of the rabbit task. One was that when watching the short-rabbit test event, infants never visually explored the screen as a whole, even when the rabbit moved out of view and the screen was the only object before them; they attended only to the thin segment of the screen at the same level as the short rabbit's face, and they never raised their eyes to explore the screen more fully. The other assumption was that even though infants saw the tall- and short-rabbit test events on alternate trials for three pairs of trials, their detection of the window in the tall-rabbit trials had no effect on their responses in the short-rabbit trials. In each trial, infants processed only a thin segment of the screen (higher in the tall-rabbit trials, lower in the short-rabbit trials), and they looked longer whenever this segment had a missing portion relative to the pretest trials.

From a cognitive standpoint, both of these assumptions are doubtful. First, it is implausible that infants failed to detect the window in the short-rabbit test event. The test screen differed from the pretest screen in both color and shape, and its window was about seven times larger than the rabbit's face; infants most likely detected the salient changes to the screen from the first test trial they received, irrespective of whether they saw the tall- or the short-rabbit event. Second, it is also implausible that infants detected the window in the tall-rabbit trials but behaved as though they were unaware of its presence in the short-rabbit trials. As was discussed in the last section, stable, detailed representations of objects are necessary for infants to successfully recognize objects in everyday life; anything less would impair cognitive functioning and learning.

9.2.2. Domain-General Account

In a VOE false-belief task (Onishi & Baillargeon, 2005), 15-month-olds first received three familiarization trials. In the first, an agent opened doors at the back of a puppet-stage apparatus, hid a watermelon toy in a green as opposed to a yellow box, and paused while holding her toy. The other two trials were identical except that the agent simply reached into the green box and held the toy. Next, in the belief-induction trial, the toy moved by itself into the yellow box, and the agent either watched this event through a window in the back wall (agent-present condition) or was absent from the scene (agent-absent condition). Finally, in the test trial, the agent entered the apparatus and reached into either the green box (green-box trial) or the yellow box (yellow-box trial). Infants in the agent-present condition looked significantly longer if they received the green-box as opposed to the yellow-box trial, whereas those in the agent-absent condition showed the reverse looking pattern. The original interpretation of these findings was that infants expected the agent to act on the information available to her, whether true or false, and they were surprised when she did not. An alternative interpretation was proposed by the *domain-general* account, which holds that low-level domain-general processes are solely responsible for infants' responses in VOE tasks (Heyes, 2014). To explain the findings described above, two processes were invoked. One was *recency-based novelty*: In the agent-present condition, infants looked longer in the green-box trial because the agent's movement toward the green box was perceptually more novel relative to the toy's movement toward the yellow box in the belief-induction trial. The other process was *retroactive interference*: In the agent-absent condition, infants looked longer in the yellow-box trial because (a) the agent's return at the start of the trial disrupted their memory of the belief-induction trial and (b) the agent's movement toward the yellow box was perceptually more novel relative to her movement toward the green box in the trial before the (now erased) belief-induction trial, the last familiarization trial.

This account rests on two assumptions that are both cognitively implausible. First, consider the assumption that the toy's movement toward the yellow box in the belief-induction trial *superseded* the agent's repeated movements toward the green box in the familiarization trials. The notion that perceptual novelty in a series of trials might be computed relative to only the *immediately preceding* trial flies in the face of much infancy research. In particular, this would mean that in a VOE or habituation task with multiple pretest and test trials, perceptual novelty would be established on a trial-by-trial basis, taking into account only the previous trial. In a preference task, for example, infants who saw the agent reach for unpreferred object-B in the first test trial and for preferred object-A in the second test trial would view both trials as perceptually novel: The first would be novel relative to the last familiarization trial, and the second would be novel relative to the first test trial. These predictions, however, are contradicted by a great deal of empirical evidence, as we have seen in this article.

Second, consider the assumption that the return of the agent at the start of the test trial in the agent-absent condition was sufficient to *erase* infants' memory of the belief-induction trial, causing them to revert back to the last familiarization trial when assessing the perceptual novelty of the test trial. The notion that the agent's arrival would constitute a "salient distractor" (Heyes, 2014, p. 650) capable of causing substantial memory disruption is implausible. First, infants were exposed to the agent's arrival throughout the testing session. Each familiarization and test trial began the same way: After the curtain was lifted at the front of the apparatus, the agent opened doors at the back of the apparatus. It thus seems unlikely that infants would have found the arrival of the agent at the start of the test trial distracting, as this was the fourth time they had witnessed it. Second, although it is true that the agent was entirely absent in the belief-induction trial, many VOE tasks have agents who come and go across trials, with little negative impact. For example, in preference tasks, the agent is typically absent in the display trial (when the positions of object-A and object-B are swapped) and returns in the test trials (Daum et al., 2012; Woodward, 1998). Similarly, in morality tasks, the helper and the hinderer often appear in alternate pretest trials, as do the fair and the unfair distributor (Hamlin et al., 2007; Meristo & Surian, 2013; see also Surian & Margoni, 2020a); if each character's absence in one trial and return in another trial was sufficient to cause memory disruption, we might expect infants to become hopelessly confused across trials, yet this is not what usually happens.

9.2.3. Hyper-Local Information Processors

Both the thin-segment and domain-general accounts portray infants as shallow, hyper-local information processors. In the thin-segment account, infants focus on restricted portions of scenes, as though perceptual blinders prevent them from seeing other portions, and they do little more than superficially compare these attended portions across events. In the domain-general account, new events can entirely supersede previous events, limiting infants to hyper-local and superficial assessments of perceptual novelty. Moreover, "salient distractors" can erase infants' memory of entire events. Such a characterization of infant cognition is at odds with what developmental psychology has revealed about how infants process the world around them.

As these examples illustrate, it becomes possible to generate numerous low-level accounts for VOE findings if one is willing to forego cognitive plausibility. VOE findings, like other findings pointing to early cognitive competencies, will often be open to alternative low-level accounts, and it is crucial, as was discussed in section 8.1, that these be addressed via appropriate control conditions. However, there usually will be only a limited number of plausible alternatives to consider, as there are only so many ways infants can plausibly process the events they are shown.

9.3. Unease at Odd Events

Another criticism of the VOE paradigm is that (a) tasks often involve events that deviate from everyday life, and (b) when watching these odd events, infants most likely experience little more than a vague sense of unease, with no real inkling about what makes the events odd (Haith, 1998). Such reactions are said to be too shallow to warrant the attribution of sophisticated expectations or reasoning abilities, casting doubt on the rich conclusions commonly drawn from such tasks. There are several difficulties with this criticism.

9.3.1. Truly Odd Events

To start, let us consider VOE tasks using events that could truly be described as odd, such as events in which objects magically disappear or change color. Three sets of findings contradict the notion that infants experience only vague unease when shown such events and do not analyze them in any way.

The first set involves situations in which infants view an odd event as surprising in one context but not another due to *conceptual variation*: Conceptual differences allow infants to spontaneously generate an explanation for the event in the latter but not the former context. Recall, for example, that 6-month-olds were surprised when a box disappeared from behind one of two adjacent screens if they knew the box to be inert, but not if they knew it to be self-propelled (Luo et al., 2009). The second set involves situations in which infants who normally show surprise at an odd event no longer do so following a *hint manipulation* designed to suggest how the event was produced. For example, recall that 5.5-month-olds were surprised when a tall rabbit failed to appear in a screen’s high window, but not if they first received a hint that two tall rabbits might be present (Baillargeon & Graber, 1987). Finally, the third set involves situations in which infants fail to show surprise at an odd event because they have not yet developed the ability to do so, yet succeed following a *priming manipulation* designed to enhance their reasoning about the event. For example, after a purple toy was hidden in a container too small to hold more than one toy, 12-month-olds failed to be surprised when an orange toy was next retrieved; however, they did detect this change violation if they were first primed, via perceptual or linguistic manipulations, to attend to the toy's color (Lin et al., 2021).

If infants merely experienced unease when shown odd events and did not analyze them in any way, it would be difficult to understand why the very same events would trigger unease under some circumstances but not others.

9.3.2. Artificial Agents

Next, let us consider VOE tasks in which infants are presented with novel non-human artificial agents such as animated circles, ovals, triangles, or squares—all agents infants never observe in the real world (Gergely et al., 1995; Johnson et al., 2010; Liu et al., 2017; Margoni et al., 2018; Meristo & Surian, 2013; Thomsen et al., 2011; see also Kominsky et al., 2022). As we have seen throughout this article, infants have little difficulty reasoning about the actions of such agents, in either psychological- or sociomoral reasoning tasks. Infants' success at distinguishing between expected and unexpected actions for such agents attests to their remarkable ability to analyze events abstractly, bringing to bear relevant aspects of their world model.

9.3.3. Everyday Events

Until now, we have focused on VOE tasks that might arguably be described as using odd events that deviate from everyday life (e.g., objects that magically disappear, animated circles that interact with each other). However, many VOE tasks do not use such events at all. Quite often, infants are presented with simple, everyday events that become unexpected only when analyzed in light of preceding modulation events. For example, consider tasks in which a human agent is seen to approach one of two objects, one of two hiding locations, or one of two other human agents (Bian & Baillargeon, 2022; Onishi & Baillargeon, 2005; Woodward, 1998). Such events do not deviate from everyday life; they become odd only when preceding events render them inconsistent with infants' world model. In addition, recall that in some VOE tasks infants view as unexpected physically possible, everyday events that happen to violate faulty rules (committing what we called errors of commission; Baillargeon & DeJong, 2017; Luo & Baillargeon, 2005b; Wang et al., 2016). If all that infants could experience was unease at odd events, it would be difficult to explain why they would ever show surprise at everyday events.

Finally, all of the evidence that has accrued about infants' ability to analyze events—for example, to integrate novel information as it becomes available via inference, to actively anticipate outcomes, and to search for explanations of (real or perceived) violations through EBL and other targeted exploratory activities—contradicts the notion that infants in VOE tasks do little more than feel vague unease at odd events.

9.4. Action Tasks as Litmus Tests

Another criticism of the VOE paradigm comes from researchers who, for one reason or another, regard action tasks as litmus tests of infants' cognitive abilities (Buttelmann et al., 2009; Mueller & Overton, 1998; Tafreshi et al., 2014; see also Engelmann & Tomasello, 2019, and the commentary by Surian & Margoni, 2020b). Compared to action tasks, VOE tasks are said to require longer and more equivocal inferential chains from behavior to cognition and, as such, to be less intuitive or conclusive. As a consequence, VOE findings remain suspect until they are confirmed by action tasks (sometimes even then), and any discrepancies that arise between VOE and action findings are taken to invalidate the VOE findings.

One difficulty with this criticism is that VOE tasks now use a wide variety of measures, including action measures (e.g., event selection, exploration of violation objects, and actions that reflect negative evaluations of violation agents), so that drawing a sharp line between VOE and action tasks is less warranted than it once was. Another difficulty is that long inferential chains from behavior to cognition often stem more from the complex scenarios shown than from the specific measures used to assess infants' understanding of these scenarios (e.g., whether infants preferentially approach a helper over a hinderer or are surprised when the target of these actions fails to do the same, they must in either case be able to correctly interpret and evaluate the actions of the helper and the hinderer; Hamlin et al., 2007).

Yet another, weightier difficulty is that the simple-minded opposition of VOE and action tasks overlooks all that can be gained from constructive efforts to integrate their findings, when they happen to differ. As noted earlier, VOE and action tasks often do converge (Buttelmann et al., 2009; Egyed et al., 2013; Feigenson & Carey, 2003; Hespos & Baillargeon, 2006; Liszkowski et al., 2008; van de Walle et al., 2000; Wang & Kohne, 2007; Xu & Baker, 2005), but much can be learned from situations where they do not.

To start, consider situations in which infants reveal particular knowledge or abilities in their VOE looking-time responses but not yet in their actions. The earliest success at integrating VOE and action findings dates back to the debate over object permanence. When positive VOE reports first appeared, proponents of action tasks argued that because young infants failed to search for hidden objects, their VOE responses could not reveal *true* object permanence but only weak precursors (Munakata et al., 1997; Shinskey et al., 2000). A heated debate ensued that, in time, tapered down to the following question: Why did young infants fail at search tasks if they could (a) represent hidden objects (as shown by VOE tasks) and (b) perform means-end actions to retrieve visible objects (as shown by action tasks)? The account that eventually carried the day was that success at search tasks requires managing not only the *separate demands* imposed by representing hidden objects and performing means-end actions but also the *total concurrent demands* imposed by these two processes (Boudreau & Bushnell, 2000; Keen & Berthier, 2004). Simply put, young infants' limited information-processing resources allow them to manage either process alone but not both together.

The integration into a single account of positive VOE object-permanence findings and negative search findings yielded important new insights into early cognitive development and the range of factors that can affect it. Today, other discrepancies are attracting experimental attention. For example, if toddlers expect individuals to divide resources fairly in VOE tasks, why do they typically act selfishly in distributive situations (Blake et al., 2014)? Similarly, if toddlers expect agents to act on their false beliefs in VOE tasks, why do they fail at traditional, elicited-prediction tasks (Setoh et al., 2016)? In each case, constructive efforts are under way to integrate the discrepant findings from these different tasks and to advance our understanding of how early expectations eventually translate into actions or elicited predictions.

Now, let us turn to converse situations in which action tasks yield positive findings but VOE tasks do not. These situations often have to do with social actions that are not morally required. For example, although infants may spontaneously help strangers in need of instrumental assistance (Warneken & Tomasello, 2006, 2007), they hold no particular expectation about whether others will do so: They look equally whether a protagonist chooses to help, ignore, or hinder an individual who is not clearly identified as one of the protagonists' ingroup members (Hamlin, 2015; Jin & Baillargeon, 2017; Lee et al., 2015; Ting et al., 2019).

Keeping track of situations in which infants choose to act prosocially toward strangers—to help them, comfort them, or punish wrongdoers who harm them—but hold no expectations about whether others will do so is crucial for achieving not only an accurate description of their world model but also a better understanding of the motivations for their prosocial actions.

Our discussion of the various ways in which discrepancies can arise between VOE and action tasks makes clear that championing one type of task at the expense of the other can only impede scientific progress. Individual researchers may prefer to use one type of task in their own work, and they are certainly entitled to do so; but no one should lose sight of the fact that both types of tasks have a role to play in uncovering the complex developmental links between infants' expectations and actions.

9.5. A Narrow Focus on Infants' Expectations

Another criticism of the VOE paradigm is that, because of its narrow focus on infants' expectations, it tends to promote an implausible view of early cognition in which expectations become the sole, exclusive factor considered when explaining infants' responses (Mueller & Overton, 1998; Tafreshi et al., 2014).

Although it is true that VOE tasks are used primarily to uncover infants' expectations, this does not mean that researchers ignore or discount the contributions of other factors to infants' responses. For example, we have seen that infants who possess the expectation necessary to detect a violation may nonetheless fail to do so due to a whole host of factors including: limitations in their perceptual abilities, information-processing capacity, or working memory; shallow problem-solving strategies that result in perseverative responding; contamination effects from prior testing experiences; and anxiety due to temperamental tendencies or attachment-related experiences. This partial list makes clear that VOE researchers are well aware that it is not possible to fully understand infants' responses by focusing on their expectations alone.

In addition, we have discussed contributions from infants' actions in several sections of this article. For example, we have seen that infants can often demonstrate the same knowledge in both VOE and action tasks; that infants who detect a violation in an event may manually choose to watch that event again; that infants faced with real or perceived physical violations may engage in targeted exploratory actions that can lead to the elaboration or revision of their world model; and that infants faced with social violations may actively avoid or punish the agents responsible for these violations.

There are at least two other ways, not yet mentioned, in which infants' actions have been found to contribute to their VOE responses, and we take advantage of this discussion to mention them. First, infants' actions can enable them to detect fine-grained violations that they otherwise would not yet detect by allowing them, incidentally, to include the necessary featural information in their event representations. For example, prior to 11.5 months, infants typically do not detect a violation if a green ball changes into a red ball when passing behind a screen too narrow to hide two balls (Wilcox, 1999). However, 10.5-month-olds detected this violation if they first handled both balls, one at a time, before being escorted to the testing room (Wilcox et al., 2007). Similarly, prior to 7.5 months, infants do not detect a violation if a dotted ball changes into a striped ball behind a narrow screen (Wilcox, 1999). However, 6.5-month-olds detected this violation if they first handled each ball, and 5.5-month-olds did too as long as they had postural support while handling each ball (Woods & Wilcox, 2013). In each case, infants' multisensory exploration resulted in rich, well-integrated information about each ball; when later retrieved during the VOE task ("I know that dotted ball … and that striped one too!"), the information enabled infants to detect the violation they were shown.

Second, infants' actions can help clarify events that seem to them ambiguous. For example, although 3-month-olds succeed at a preference task when shown events they can readily interpret (e.g., an agent visible from the waist up reaching for and grasping objects; Choi et al., 2018), they fail with more minimal events involving only the agent's hand (Sommerville et al., 2005). However, if the hand wears a mitten, and infants are first given experience picking up similar (Velcro-covered) objects while wearing similar (Velcro-covered) mittens, they can then better understand the agent's actions and detect a violation if she fails to act on her preference after the objects' positions are swapped (Gerson & Woodward, 2014; Sommerville et al., 2005). Likewise, although 3-month-olds succeed at a detour task in which an agent reaches over a barrier to grasp an object with her bare hand, they fail if she wears a mitten, making the goal of her actions harder to fathom (Liu et al., 2019; Skerry et al., 2013). However, if first given experience at picking up similar objects while wearing similar mittens, they detect a violation if the agent fails to act efficiently after the barrier is removed (Skerry et al., 2013).

In sum, infants' actions can contribute to their VOE responses in many different ways, in line with long-standing claims about the reciprocal interactions of action and cognition in infancy (Adolph et al., 2000; Bruner, 1966; Flavell, 1977; Gibson, 1950; Piaget, 1954; Rochat & Goubet, 1995; Thelen & Smith, 1994; von Hofsten, 2007).

9.6. Adultocentric Interpretations

Another criticism of the VOE paradigm is that its findings often receive conceptual interpretations that grant infants innate adult-like expectations. Such a view, it is argued, overlooks the possibility of partial, incremental developments and leaves little room for the consideration of learning mechanisms that might contribute to these developments (Fischer & Bidell, 1991; Haith, 1998; Haith & Benson, 1998; Tafreshi et al., 2014).

One difficulty with this criticism is that it fails to appreciate that crediting infants with innate expectations can still leave ample room for incremental developments. We have already seen an example of such an approach. According to the EBL account (Baillargeon & DeJong, 2017; Wang & Baillargeon, 2008; Luo et al., 2009; Lin et al., 2022), infants' physical reasoning is constrained from birth by a skeletal framework of principles (e.g., persistence, gravity) and concepts (e.g., force, internal energy) (see also Gelman, 1990; Leslie, 1995). However, because this framework can be applied only to the representations infants form for events, and these representations are initially sparse and lacking in featural detail, infants often fail to detect fine-grained violations. Infants' event representations become more detailed as they acquire, through the EBL process, rules that identify causally relevant features for each event category. Development is thus incremental and protracted as infants identify, feature by feature and event category by event category, a myriad of rules that result in more detailed event representations, making possible more accurate predictions and better adapted actions.

Another difficulty with the criticism that conceptual interpretations of VOE findings tend to grant infants innate adult-like expectations is that it fails to acknowledge the wide variety of conceptual accounts that have been proposed for particular findings. To illustrate, consider infants' responses in psychological-reasoning tasks (Baillargeon et al., 2016). According to a *statistical* account (Ruffman et al., 2012), infants gather a wealth of statistical information about the actions agents typically produce in everyday life, and they use these statistical regularities to predict agents' actions. According to a *minimalist* account (Butterfill & Apperly, 2013), psychological reasoning in the first four years of life depends on a minimal system that can track what objects and events an agent registers in a scene and can use these registrations to predict the agent's actions in the same or subsequent scenes. According to a *constructivist* account

(Woodward et al., 2001), domain-general processes of comparison, abstraction, and generalization, applied to infants' representations of their own actions and those of other agents, make possible the gradual construction of an abstract understanding of intentional action; this understanding is then extended, broadly and flexibly, to novel actions and novel agents. According to a *conceptual-change* account (Gergely & Csibra, 2003), infants in the first year of life reason about others' actions using a non-mentalistic, teleological system; after the first birthday, this system becomes incorporated into a more advanced, mentalistic system that can, for the first time, make sense of agents' actions in terms of goals, beliefs, and other mental states. Finally, according to a *mentalistic* account (Scott et al., 2022), infants are born equipped with a mentalistic psychological-reasoning system similar to that of older children and adults; this system provides infants with a skeletal framework for representing and learning about agents' mental states, and it gradually becomes more skilled and more nuanced with age and experience. Our aim here is not to adjudicate among these different conceptual accounts but simply to point out that they posit a wide range of developmental mechanisms and trajectories.

A final, more general difficulty with the criticism that VOE findings are often taken to reveal innate adult-like expectations is that it fails to distinguish between criticisms of the VOE paradigm and criticisms of particular interpretations that have been proposed for some of its findings. As Ginnobili and Olmos (2021) pointed out, the VOE paradigm "is neutral regarding how whatever generated the unfulfilled expectations should be interpreted" (p. 3); an innate-knowledge interpretation "is not an essential element of this experimental paradigm, but one of the possible interpretations of its results" (p. 4). While we were preparing this article, Paulus (2022) published an essay entitled "Should infant psychology rely on the violation-of-expectation method? Not anymore" (for a reply, see Stahl & Kibbe, 2022). In his essay, Paulus reviews many of the criticisms of the VOE paradigm we have discussed, including the notion that “a variety of findings could be explained without the need to ascribe conceptual understanding to infants” (p. 7). In offering this argument, Paulus fails to distinguish between criticisms of the VOE paradigm and criticisms of interpretations of its findings. At many times throughout the history of the paradigm, researchers have offered widely different interpretations of the same VOE findings; such disagreements are simply a part of the process of science. Rejecting the VOE paradigm because one disagrees with interpretations of its findings would be tantamount to rejecting the manual-search paradigm Piaget (1952, 1954) used to investigate the development of object permanence because one disagreed with his interpretations of his findings.

10. Conclusions

The evidence reviewed in this article suggests four broad conclusions. First, the VOE paradigm is well-suited for examining whether infants hold expectations about how particular events are likely to unfold. Second, the VOE paradigm is also well-suited for tracking when expectations emerge and change, and for exploring the various mechanisms that contribute to these developments. Third, in many VOE tasks, expectations are uncovered only under optimal circumstances, when everything about the task is carefully designed to support infants’ reasoning and to refrain from overtaxing their limited information-processing resources. In practical terms, the reasoning infants reveal in VOE tasks is typically far removed from the robust, skilled, and efficient reasoning older children demonstrate in extensive batteries of tasks. Nevertheless, knowing what specific components and precursors of this reasoning are already in place in infancy can be critical for constraining theorizing about developmental mechanisms and trajectories. Finally, the VOE paradigm has changed considerably since its inception; not only has its conceptual rationale evolved, but significant extensions in its measures have helped broaden and strengthen it. Happily, the many different measures used in VOE tasks today paint a coherent picture of how infants form expectations about events’ outcomes, are surprised when these expectations are violated, and seek explanations that reconcile their observations with their working model of the world. With the recent calls for better research practices and for more sophisticated data analyses in psychological research, the VOE paradigm is certain to change again, and for the better. Still, at its core, it will remain what it has always been: a valuable tool for exploring the infant mind and its development.

Acknowledgments

The preparation of this manuscript was supported by a John Templeton Foundation grant (No. 60847) to R. Baillargeon. We thank three anonymous reviewers as well as Jerry DeJong, Cindy Fisher, Erik Fonn, Kyong-sun Jin, Lotte Thomsen, Fei Xu, and Joakim Zahl for helpful comments and suggestions.

References

Adolph, K. E., Eppler, M. A., Marin, L., Weise, I. B., & Clearfield, M. W. (2000). Exploration in the service of prospective control. *Infant Behavior and Development: Special Issue on Perception – Action Coupling, 23,* 441-460.

Aguiar, A., & Baillargeon, R. (1999). 2.5-month-old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology, 39,* 116-157.

Aguiar, A., & Baillargeon, R. (2002). Developments in young infants' reasoning about occluded objects. *Cognitive Psychology, 45,* 267-336.

Aguiar, A., & Baillargeon, R. (2003). Perseverative responding in a violation-of-expectation task in 6.5-month-old infants. *Cognition, 88,* 277-316.

Aslin, R. N. (2007). What's in a look? *Developmental Science, 10,* 48-53.

Aslin, R. N., & Fiser, J. (2005). Methodological challenges for understanding cognitive development in infants. *Trends in Cognitive Sciences, 9,* 92-98.

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition, 23,* 21-41.

Baillargeon, R. (1987). Object permanence in 3.5 and 4.5-month-old infants. *Developmental Psychology, 23,* 655-664.

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science, 3,* 133-140.

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, NJ: Ablex.

Baillargeon, R. (2002). The acquisition of physical knowledge in infancy: A summary in eight lessons. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 47-83). Oxford: Blackwell.

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science, 13*, 89-94.

Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Invited commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development, 46,* 112-124.

Baillargeon, R., & DeJong, G. F. (2017). Explanation-based learning in infancy. *Psychonomic Bulletin & Review, 24,* 1511-1526.

Baillargeon, R., & DeVos, J. (1991). Object permanence in 3.5- and 4.5-month-old infants: Further evidence. *Child Development, 62,* 1227-1246.

Baillargeon, R., & Graber, M. (1987). Where's the rabbit? 5.5-month-old infants' representation of the height of a hidden object. *Cognitive Development, 2,* 375-392.

Baillargeon, R., Li, J., Ng, W., & Yuan, S. (2009). An account of infants physical reasoning. In A. Woodward & A. Neeham (Ed.), *Learning and the infant mind* (pp. 66-116). Oxford: 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.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67*, 159-186.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in 5-month-old infants. *Cognition, 20,* 191-208.

Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., Bolz, M., Henrich, J., Setoh, P., Wang, J., & Laurence, S. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B, 280*, 20122654.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science, 15,* 611-617.

Berger, A., Tzur, G., & Posner, M. I. (2006). Infant brains detect arithmetic errors. *Proceedings of the National Academy of Sciences, USA, 103,* 12649-12653.

Bian, L., & Baillargeon, R. (2022). When are similar individuals a group? Early reasoning about similarity and in-group support. *Psychological Science, 33*, 752-764.

Bian, L., Sloane, S., & Baillargeon, R. (2018). Infants expect ingroup support to override fairness when resources are limited. *Proceedings of the National Academy of Sciences, USA**,** 115,* 2705-2710.

Bird, C. D., & Emery, N. J. (2010). Rooks perceive support relations similar to six-month-old babies. *Proceedings of the Royal Society B, 277*, 147–151.

Bíró, S., Verschoor, S., & Coenen, L. (2011). Evidence for a unitary goal concept in 12-month-old infants. *Developmental Science, 14,* 1255-1260.

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge-behavior gap. *Trends in Cognitive Sciences, 18*, 559-561.

Bogartz, R. S., Shinskey, J. L., & Schilling, T. H. (2000). Object permanence in five-and-a half-month-old infants? *Infancy, 1,* 403-428.

Bogartz, R. S., Shinskey, J. L., & Speaker, C. J. (1997). Interpreting infant looking: The event set x event set design. *Developmental Psychology, 33,* 408-422.

Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A* *methodological overview* (pp. 253-300). Norwood, NJ: Ablex.

Boudreau, J. P., & Bushnell, E. W. (2000). Spilling thoughts: Configuring attentional resources in infants' goal-directed actions. *Infant Behavior and Development*, *23,* 543-566.

Bower, T. G. R. (1974). *Development in infancy*. San Francisco, CA: W. H. Freeman.

Bower, T. G. R., Broughton, J. M., & Moore, M. K. (1971). Development of the object concept as manifested in the tracking behavior of infants between 7 and 20 weeks of age. *Journal of Experimental Child Psychology, 11,* 182-193.

Brandone, A. C., Horwitz, S. R., Aslin, R. N., & Wellman, H. M. (2014). Infants' goal anticipation during failed and successful reaching actions. *Developmental Science, 17,* 23-34.

Brooker, I., & Poulin-Dubois, D. (2013). Is a bird an apple? The effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy, 18,* 46-68.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition, 104,* 287-314.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112,* 337-342.

Butterfill, S., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind and Language, 28,* 606-637.

Buyukozer Dawkins, M., Sloane, S., & Baillargeon, R. (2019). Do infants in the first year of life expect equal resource allocations? *Frontiers in Psychology, 10,* 116.

Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). Six solutions for more reliable infant research. *Infant and Child Development*, e2296.

Cacchione, T., & Krist, H. (2004). Recognizing impossible object relations: Intuitions about support in chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *118,* 140-148.

Cacchione, T., Schaub, S., & Rakoczy, H. (2013). Fourteen-month-old infants infer the continuous identity of objects on the basis of nonvisible causal properties. *Developmental Psychology, 49*, 1325-1329.

Camras, L. A., Meng, Z., Ujiie, T., Dharamsi, S., Miyake, K., Oster, H., Wang, L., Cruz, J., Murdoch, A., & Campos, J. (2002). Observing emotion in infants: Facial expression, body behavior, and rater judgments of responses to an expectancy-violating event. *Emotion, 2*, 179-193.

Cannon, E., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, *15,* 292-98.

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

Carpenter, M., Nagell, K., Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*, 1-143.

Cashon, C. H., & Cohen, L. B. (2000). Eight-month-old infants' perception of possible and impossible events. *Infancy, 1,* 429-446.

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. L. (2018). Precursors of logical reasoning in preverbal human infants. *Science, 359,* 1263-1266.

Charlesworth, W. R. (1969). The role of surprise in cognitive development. In D. Elkind & J. H. Flavell (Eds.), *Studies in cognitive development: Essays in honor of Jean Piaget* (pp. 257-314). New York: Oxford University Press.

Choi, Y., & Luo, Y. (2015). 13-month-olds' understanding of social interactions. *Psychological Science*, *26,* 274-283.

Choi, Y., Luo. Y., & Baillargeon, R. (2022). Can 5-month-old infants consider the perspective of a novel eyeless agent? New evidence for early mentalistic reasoning. *Child Development, 93,* 571-581.

Choi, Y. J., Mou, Y., & Luo, Y. (2018). How do 3-month-old infants attribute preferences to a human agent? *Journal of Experimental Child Psychology*, *172*, 96-106.

Chow, V., Poulin-Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science, 11*, 761-770.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*, 181-204.

Cohen, L. B. (1976). Habituation of infant visual attention. In T. J. Tighe & R. N. Leaton (Eds.) *Habituation: Perspectives from child development, animal behavior, and neurophysiology* (pp. 207–238). Hillsdale, NJ: Lawrence Erlbaum.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*, 705-717.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27,* 111-133.

Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of "pure reason" in infancy. *Cognition, 72,* 237-267.

Dasgupta, A., Duan, J., Lin, Y., Wang, S., Baillargeon, R., & Tan, C. (2023). A benchmark for modeling violation-of-expectation in physical reasoning across event categories. In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th*

Annual Conference of the Cognitive Science Society, 45, 1755-1762.

Daum, M., Attig, M., Gunawan, R., Prinz, W., & Gredebäck, G. (2012). Actions seen through babies' eyes: A dissociation between looking time and predictive gaze. *Frontiers in Psychology, 3,* 370.

Davis-Kean, P. E., & Ellis, A. (2019). An overview of issues in infant and developmental research for the creation of robust and replicable science. *Infant Behavior and Development, 57,* 101339.

DeBolt, M. C., Rhemtulla, M., & Oakes, L. M. (2020). Robust data and power in infant research: A case study of the effect of number of infants and number of trials in visual preference procedures. *Infancy, 25*, 393-419.

DesChamps, T. D., Eason, A. E., & Sommerville, J. A. (2016). Infants associate praise and admonishment with fair and unfair individuals. *Infancy, 21*, 478-504.

Dewar, K., & Xu, F. (2007). Do 9-month-old infants expect distinct words to refer to kinds? *Developmental Psychology, 43,* 1227.

Drayton L. A., & Santos, L. R. (2018). What do monkeys know about others' knowledge? *Cognition, 170,* 201-208.

Dykas, M. J., & Cassidy, J. (2011). Attachment and the processing of social information across the life span: Theory and evidence. *Psychological Bulletin, 137*, 19-46.

Eason, A. E., Hamlin, J. K., & Sommerville, J. A. (2017). A survey of common practices in infancy research: Description of policies, consistency across and within labs, and suggestions for improvements. *Infancy, 22,* 470-491.

Egyed, K., Király, I., & Gergely, G. (2013). Communicating shared knowledge in infancy. *Psychological Science, 24*, 1348-1353.

Engelmann, J. M., & Tomasello, M. (2019). Children's sense of fairness as equal respect. *Trends in Cognitive Sciences, 23,* 454-463.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146,* 668-670.

Farmer, T., Brown, M., & Tanenhaus, M. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences, 36*, 211-212.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44,* 491-505.

Feigenson, L. & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science, 6,* 568-584.

Fischer, K. W., & Bidell, T. (1991). Constraining nativist inferences about cognitive capacities. In S. Carey & R. Gelman (Eds.). *The epigenesis of mind: Essays on biology and cognition* (pp. 199-235). Hillsdale, NJ: Erlbaum.

Flavell, J. H. (1977). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.

Forgács, B., Gervain, J., Parise, E., Csibra, G., Gergely, G., Baross, J., & Király, I. (2020), Electrophysiological investigation of infants' understanding of understanding. *Developmental Cognitive Neuroscience, 43,* 100783.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., … Lew-Williams, C. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy, 22,* 421-435.

Gehring, W. J., Coles, M. G. H., Meyer, D. E., & Donchin, E (1995). A brain potential manifestation of error-related processing. In G. Karmos, M. Molnár, V. Csépe, I. Czigler, & J. E. Desmedt (Eds.), *Perspectives of event-related potentials research* (EEG Suppl. 44, pp. 261-272). Amsterdam: Elsevier.

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

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants’ reactions to equal and unequal distributions of resources. *Developmental Science, 14,* 1012-1020.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7,* 287-292.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56,* 165-193.

Gerson, S. A., & Woodward, A. L. (2014). The joint role of trained, untrained, and observed actions at the origins of goal recognition. *Infant Behavior and Development*, *37*, 94-104.

Gibson, J. J. (1950). *The perception of the visual world*. Boston, MA: Houghton-Mifflin.

Gilfillan, G., Vitale, J., McNutt, J. W., & McComb, K. (2016). Cross-modal individual recognition in wild African lions. *Biology Letters, 12,* 20160323.

Ginnobili, S., & Olmos, A. S. (2021). Empirical assumptions behind the violation of expectation experiments in human and non-human animals. *History and Philosophy of the Life Sciences, 43,* 106.

Gredebäck, G., Johnson, S., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology, 35*, 1-19.

Gredebäck, G., Lindskog, M., Juvrud, J. C., Green, D., & Marciszko, C. (2018). Action prediction allows hypothesis testing via internal forward models at 6 months of age. *Frontiers in Psychology, 9,* 290.

Gredebäck, G., & Melinder, A. (2010). Infants’ understanding of everyday social interactions: A dual process account. *Cognition, 114*, 197-206.

Grosse Wiesmann, C. G., Friederici, A. D., Disla, D., Steinbeis, N., & Singer, T. (2018). Longitudinal evidence for 4-year-olds' but not 2-and 3-year-olds' false belief-related action anticipation. *Cognitive Development*, *46*, 58-68.

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development, 21,* 167-179.

Haith, M. M., & Benson, J. B. (1998). Infant cognition. In W. Damon (Series Ed.) & D. Kuhn & R. Siegler (Eds), *Handbook of child psychology* (Vol. 2, pp. 199-254). New York: Wiley.

Hamlin, J. K. (2015). The case for social evaluation in preverbal infants: Gazing toward one's goal drives infants' preferences for Helpers over Hinderers in the hill paradigm. *Frontiers in Psychology*, 5, 1563.

Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as I do: 7-month-old infants selectively reproduce others' goals. *Developmental Science, 11,* 487–494.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450,* 557-559.

Hepach, R., & Westermann, G. (2016). Pupillometry in infancy research. *Journal of Cognition and Development, 17,* 359-377.

Hespos, S. J., & Baillargeon, R. (2001). Infants' knowledge about occlusion and containment events: A surprising discrepancy. *Psychological Science*, *12*, 141-147.

Hespos, S. J., & Baillargeon, R. (2006). Décalage in infants' knowledge about occlusion and containment events: Converging evidence from action tasks. *Cognition, 99*, B31-B41.

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.

Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science*, *17,* 647-659.

Hohenberger, A., Elsabbagh, M., Serres, J., de Schoenen, S., Karmiloff-Smith, A., & Aschersleben, G. (2012). Understanding goal-directed human actions and physical causality: The role of mother-infant interaction. *Infant Behavior and Development*, *35*, 898-911.

Huettel, S. A., & Needham, A. (2000). Effects of balance relations between objects on infants' object segregation. *Developmental Science, 3,* 415-427.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 5, pp. 69-95). Norwood, NJ: Ablex.

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel or familiar stimuli. *Developmental Psychology, 19,* 338-352.

Hunter, M. A., Ross, H. S., & Ames, E. W. (1982). Preferences for familiar or novel toys: Effects of familiarization time in 1-year-olds. *Developmental Psychology, 18,* 519-529.

Jin, K., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences, USA*, *114*, 8199-8204.

Jin, K., & Baillargeon, R. (2023). *Who can be harmed? Infants' expectations about acceptable harm within and between moral circles*. Manuscript in preparation.

Jin, K., Houston, J. L., Baillargeon, R., Groh, A. M., & Roisman, G. I. (2018). Young infants expect an unfamiliar adult to comfort a crying baby: Evidence from a standard violation-of-expectation task and a novel infant-triggered-video task. *Cognitive Psychology, 102,* 1-

20.

Johnson, S. C., Dweck, C. S., Chen, F. S., Stern, H. L., Ok, S.-J., & Barth, M. E. (2010). At the intersection of social and cognitive development: Internal working models of attachment in infancy. *Cognitive Science, 34,* 807-825.

Johnson, S. C., Shimizu, Y. A. & Ok, S.-J. (2007). Actors and actions: The role of agent behavior in infants' attribution of goals. *Cognitive Development*, *22*, 310-322.

Káldy, Z., & Leslie, A. M. (2003). Identification of objects in 9-month-old infants: integrating 'what' and 'where' information. *Developmental Science*, *6*, 360-373.

Kamewari, K., Kato, M., Kanda, T., Ishiguro, H., & Hiraki, K. (2005). Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Development*, *20,* 303-320.

Keen, R. E., & Berthier, N. E. (2004). Continuities and discontinuities in infants' representation of objects and events. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 32, pp. 243-279). San Diego, CA: Elsevier Academic Press.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One*, *7*, e36399.

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, *119,* 197-215.

Kim, E., & Song, H. (2015). Six-month-olds actively predict others' goal-directed actions. *Cognitive Development, 33,* 1-13.

Kominsky, J. F. (2022). The challenges of improving infant research methods. *Infant and Child Development,* e2332.

Kominsky, J. F., Lucca, K., Thomas, A. J., Frank, M. C., & Hamlin, J. K. (2022). Simplicity and validity in infant research. *Cognitive Development*, *63*, 101213.

Köster, M., Kayhan, E., Langeloh, M., & Hoehl, S. (2020). Making sense of the world: Infant learning from a predictive processing perspective. *Perspectives on Psychological Science, 15,* 562-571.

Köster, M., Langeloh, M., & Hoehl, S. (2019). Visually entrained theta oscillations increase for unexpected events in the infant brain. *Psychological Science, 30,* 1656-1663.

Kotovsky, L., & Baillargeon, R. (1994). Calibration-based reasoning about collision events in 11-month-old infants. *Cognition, 51*, 107-129.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*, 402-408.

Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cognitive Development, 46*, 97-111.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science, 7*, 18-27.

Lee, Y., Yun, J., Kim, E., & Song, H. (2015). The development of infants' sensitivity to behavioral intentions when inferring others' social preferences. *PLoS One*, *10*, e0135588.

Leslie, A. M. (1995). A theory of agency. In D. Sperber & D. Premack & A. J. Premack, *Causal Cognition: A Multidisciplinary Debate* (pp. 121-149). Oxford: Clarendon Press.

Leslie, A. M. (2004). Who's for learning? *Developmental Science, 7,* 417-419.

Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences, 12,* 213-218.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25,* 265-288.

Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing ‘what’ and ‘where’ systems. *Trends in Cognitive Sciences, 2,* 10-18.

Lewry, C., Curtis, K., Vasilyeva, N., Xu, F., & Griffiths, T. L. (2021). Intuitions about magic track the development of intuitive physics. *Cognition*, *214*, 104762.

Lin, Y., Li, J., Gertner, Y., Ng, W., Fisher, C. L., & Baillargeon, R. (2021). How do the object-file and physical-reasoning systems interact? Evidence from priming effects with object arrays or novel labels. *Cognitive Psychology, 125,* 101368.

Lin, Y., Stavans, M., & Baillargeon, R. (2022). Infants’ physical reasoning and the cognitive architecture that supports it. In O. Houdé & G. Borst (Eds.), *Cambridge handbook of cognitive development* (pp. 168-194). Cambridge, England: Cambridge University Press.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition, 108,* 732-739.

Liu, S., Brooks, N. B., & Spelke, E. S. (2019). Origins of the concepts cause, cost, and goal in prereaching infants. *Proceedings of the National Academy of Sciences*, *USA*, *116*, 17747-17752.

Liu, S., Lydic, K., & Saxe, R. (2023). Using fMRI to study the neural basis of violation-of-expectation. *Journal of Vision, 23,* 4925.

Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science, 358*, 1038-1041.

Low, J., & Edwards, K. (2018). The curious case of adults' interpretations of violation-of-expectation false belief scenarios. *Cognitive Development, 46*, 86-96.

Luo, Y., & Baillargeon, R. (2005a). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science, 16,* 601-608.

Luo, Y., & Baillargeon, R. (2005b). When the ordinary seems unexpected: Evidence for incremental physical knowledge in young infants. *Cognition, 95,* 297-328.

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*, 489-512.

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science, 12,* 142-149.

Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology, 58*, 441-486.

Margoni, F. (2022). Reliability and replicability in infant research: A commentary on Byers-Heinlein et al. (2021). *Infant and Child Development*, *31*, e2330.

Margoni, F., Baillargeon, R., & Surian, L. (2018). Infants distinguish between leaders and bullies. *Proceedings of the National Academy of Sciences, USA, 115,* 8835-8843.

Margoni, F., & Shepperd, M. (2020). Changing the logic of replication: A case from infant studies. *Infant Behavior and Development, 61,* 101483.

Margoni, F., & Surian, L. (2018). Infants' evaluation of prosocial and antisocial agents: A meta-analysis. *Developmental Psychology, 54*, 1445-1455.

Margoni, F., Surian, L., Hadjichristidis, C., & Thomsen, L. (2023). *Infants show vigilance* *toward unfair distributors*. Manuscript in preparation.

McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science, 15*, 776-781.

Meicler, M., & Gratch, G. (1980). Do 5-month-olds show object conception in Piaget's sense? *Infant Behavior and Development, 3,* 265-282.

Meristo, M., Strid, K., & Surian, L. (2016). Preverbal infants' ability to encode the outcome of distributive actions. *Infancy, 21,* 353-372.

Meristo, M., & Surian, L. (2013). Do infants detect indirect reciprocity? *Cognition, 129,* 102-113.

Meristo, M., & Surian, L. (2014). Infants distinguish antisocial actions directed towards fair and unfair agents. *PloS one, 9,* e110553.

Moore, M. K., Borton, R., & Darby, B. L. (1978). Visual tracking in young infants: Evidence for object identity or object permanence? *Journal of Experimental Child Psychology, 25,* 183-198.

Mueller, U., & Overton, W. F. (1998). How to grow a baby: A re-evaluation of image-schema and Piagetian action approaches to representation. *Human Development, 41,* 71-111.

Muller, A. A., & Aslin, R. N. (1978). Visual tracking as an index of the object concept. *Infant Behavior and Development, 1*, 309-319.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104,* 686-713.

Murai, C., Tanaka, M., & Sakagami, M. (2011). Physical intuitions about support relations in monkeys (Macaca fuscata) and apes (Pan troglodytes). *Journal of Comparative*

Psychology, 125, 216-226.

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

Newcombe, N., Huttenlocher, J., & Learmonth, A. (1999). Infants' coding of location in continuous space. *Infant Behavior and Development*, *22*, 483-510.

Newcombe, N. S., Sluzenski, J., & Huttenlocher, J. (2005). Preexisting knowledge versus on-line learning: What do young infants really know about spatial location? *Psychological Science*, *16*, 222-227.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255-258.

Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica, 124*, 106-128.

OSF. (2023). *The violation-of-expectation paradigm: A conceptual overview*. Open Science Framework. https://osf.io/k3pae/

Parise, E., Csibra, G., (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, *23*, 728-733.

Pätzold, W., & Liszkowski, U. (2019). Pupillometry reveals communication-induced object expectations in 12- but not 8-months-old infants. *Developmental Science, 22,* e12832.

Paulus, M. (2022). Should infant psychology rely on the violation-of-expectation method? Not anymore. *Infant and Child Development, 31,* e2306.

Perez, J., & Feigenson, L. (2021). Stable individual differences in infants' responses to violations of intuitive physics. *Proceedings of the National Academy of Sciences, USA, 118,* e2103805118.

Perez J., & Feigenson, L. (2022). Violations of expectation trigger infants to search for explanations. *Cognition, 218*, 104942.

Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition*, *98*, 137-155.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic.

Piloto, L.S., Weinstein, A., Battaglia, P., & Botvinick. M. (2022). Intuitive physics learning in a deep-learning model inspired by developmental psychology. *Nature Human Behaviour*, *6*, 1257–1267.

Poulin-Dubois, D., & Chow, V. (2009). The effect of a looker’s past reliability on infants’ reasoning about beliefs. *Developmental Psychology, 45*, 1576-1582.

Riochet, R., Castro, M., Bernard, M., Lerer, A., Fergus, R., Izard, V., & Dupoux, E. (2018). Intphys: A framework and benchmark for visual intuitive physics reasoning. Preprint at https://arxiv.org/abs/1803.07616.

Rochat, P., & Goubet, N. (1995). Development of sitting and reaching in 5- to 6-month-old infants. *Infant Behavior and Development, 18*, 53-68.

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology, 18,* 704-713.

Rubio-Fernández, P. (2019). Publication standards in infancy research: Three ways to make violation-of-expectation studies more reliable. *Infant Behavior and Development, 54,* 177-188.

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology, 30*, 87-104.

Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who-dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology, 43*, 149-158.

Scherer, K. R., Zentner, M. R., & Stern, D. (2004). Beyond surprise: The puzzle of infants' expressive reactions to expectancy violation. *Emotion, 4,* 389-402.

Schuwerk, T., et al. (2022). Action anticipation based on an agent's epistemic state in toddlers and adults. *Child Development*, In-Principle Acceptance of Registered Report Stage 1: Study Design.

Scott, R. M., He, Z., Baillargeon, R., & Cummins, D. (2012). False-belief understanding in 2.5-year-olds: Evidence from two novel verbal spontaneous-response tasks. *Developmental Science*, *15*, 181-193.

Scott, R. M., Baillargeon, R., Song, H. J., & Leslie, A. M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, *61*, 366-395.

Scott, R. M., Richman, J. C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology*, *82,* 32-56.

Scott, R. M., Roby, E., & Baillargeon, R. (2022). How sophisticated is infants' theory of mind? In O. Houdé & G. Borst (Eds.), *Cambridge handbook of cognitive development* (pp. 242-268). Cambridge, England: Cambridge University Press.

Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences, USA, 113*, 13360-13365.

Setoh, P., Wu, D., Baillargeon, R., & Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences, USA, 110,* 15937-15942.

Shinskey, J. L., Bogartz, R. S., & Poirier, C. R. (2000). The effects of graded occlusion on manual search and visual attention in 5- to 8-month-old infants. *Infancy, 1,* 323-346.

Sim, Z. L., & Xu, F. (2017). Infants preferentially approach and explore the unexpected. *British Journal of Developmental Psychology, 35,* 596-608.

Singer, R., & Henderson, E. (2015). Object permanence in marine mammals using the violation of expectation procedure. *Behavioural Processes, 112,* 108-113.

Skerry, A. E., Carey, S. E., & Spelke, E. S. (2013). First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proceedings of the National Academy of Sciences, USA, 110,* 18728-18733.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science, 23,* 196-204.

Smith, K. A., Mei, L, Yao, S., Wu, J., Spelke, E., et al. (2019). Modeling expectation violation in intuitive physics with coarse probabilistic object representations. *Advances in Neural Information Processing Systems, 32*.

Sodian, B., Licata, M., Kristen-Antonow, S., Paulus, M., Killen, M., &Woodward, A. (2016). Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: A longitudinal investigation. *Child Development, 87,* 1221-1232.

Sodian, B., Schoeppner, B., & Metz, U. (2004). Do infants apply the principle of rational action to human agents? *Infant Behavior and Development, 27,* 31-41.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition, 96,* B1-B11.

Southgate, V., & Begus, K. (2013). Motor activation during the prediction of nonexecutable actions in infants. *Psychological Science, 24,*828-835.

Spelke, E. S. (1985). Preferential looking methods as tools for the study of cognition in infancy. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 323-363). Norwood, NJ: Ablex.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99,* 605-632.

Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion, and object identity in infancy. *British Journal of Developmental Psychology, 13,* 1-30.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science, 348,* 91-94.

Stahl, A. E., & Kibbe, M. M. (2022). Great expectations: The construct validity of the violation-of-expectation method for studying infant cognition. *Infant and Child Development,* e2359.

Stavans, M., Lin, Y., Wu, D., & Baillargeon, R. (2019). Catastrophic individuation failures in infancy: A new model and predictions. *Psychological Review, 126,* 196-225.

Surian, L., & Franchin, L. (2017). Infants reason about deserving agents: A test with distributive actions. *Cognitive Development, 44,* 49-56.

Surian, L., & Margoni, F. (2020a). First steps toward an understanding of procedural fairness. *Developmental Science, 23,* e12939.

Surian, L., & Margoni, F. (2020b). Commentary: Children's sense of fairness as equal respect. *Frontiers in Psychology, 11,* 107.

Surian, L., Ueno, M., Itakura, S., & Meristo, M. (2018). Do infants attribute moral traits? Fourteen-month-olds' expectations of fairness are affected by agents' antisocial actions. *Frontiers in Psychology, 9,* 1649.

Tafreshi, D., Thompson, J. J., & Racine, T. P. (2014). An analysis of the conceptual foundations of the infant preferential looking paradigm. *Human Development*, 57, 222-240.

Takagi, S., Arahori, M., Chijiiwa, H., Tsuzuki, M., Hataji, Y., & Fujita, K. (2016). There's no ball without noise: Cats' prediction of an object from noise. *Animal Cognition*, *19*, 1043-1047.

Téglás, E., & Bonatti, L. L. (2016). Infants anticipate probabilistic but not deterministic outcomes. *Cognition, 157,* 227-236.

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences, USA, 104,* 19156-19159.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Thoermer, C., Woodward, A., Sodian, B., Perst, H., & Kristen, S. (2013). To get the grasp: Seven-month-olds encode and selectively reproduce goal-directed grasping. *Journal of Experimental Child Psychology, 116,* 499-509.

Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M., & Carey. S. (2011). Big and mighty: Preverbal infants mentally represent social dominance. *Science, 331,* 477-480.

Ting, F., & Baillargeon, R. (2021). Toddlers draw broad negative inferences from wrongdoers' moral violations. *Proceedings of the National Academy of Sciences, USA, 118,* e2109045118.

Ting, F., He, Z., & Baillargeon, R. (2019). Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proceedings of the National Academy of Sciences, USA, 116*, 6025-6034.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*, 383-403.

Van de Walle, G. A., Carey, S., & Prevor, M. (2000). Bases for object individuation in infancy: Evidence from manual search. *Journal of Cognition and Development, 1*, 249-280.

von Hofsten, C. (2007). Action in development. *Developmental Science, 10*, 54-60.

Völter, C. J., Tomašić, A., Nipperdey, L., & Huber, L. (2023). Dogs' expectations about occlusion events: from expectancy violation to exploration. *Proceedings of the Royal Society B, 290*, 20230696.

Wang, S. (2019). Regularity detection and explanation-based learning jointly support learning about physical events in early infancy. *Cognitive Psychology, 113*, 101219.

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.

Wang, S., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. *Cognition, 93*, 167-198.

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.

Wang, S., & Kohne, L. (2007). Visual experience enhances infants' use of task-relevant information in an action task. *Developmental Psychology, 43*, 1513-1522.

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.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*, 1301-1303.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*, 271-294.

Weihs, L., Yuile, A. R., Baillargeon, R., Fisher, C. L., Marcus, G., Mottaghi, R., & Kembhavi, A. (2022). Benchmarking progress to infant-level physical reasoning in AI. *Transactions on Machine Learning Research*, published 10/2022.

Wellman, H. (2011). Reinvigorating explanations for the study of early cognitive development. *Child Development Perspectives, 5*, 33-38.

Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition, 72*, 125–166.

Wilcox, T., & Baillargeon, R. (1998a). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology, 37,* 97-155.

Wilcox, T., & Baillargeon, R. (1998b). Object individuation in young infants: Further evidence with an event-monitoring paradigm. *Developmental Science, 1,* 127-142.

Wilcox, T., Nadel, L., & Rosser, R. (1996). Location memory in healthy preterm and full-term infants. *Infant Behavior and Development, 19,* 309-323.

Wilcox, T., Schweinle, A., & Chapa, C. (2003). Object individuation in infancy. In F. Fagan & H. Hayne (Eds.), *Progress in infancy research* (Vol. 3, pp. 193–243). Mahwah, NJ: Erlbaum.

Wilcox, T., Woods, R., Chapa, C., & McCurry, S. (2007). Multisensory exploration and object individuation in infancy. *Developmental Psychology, 43*, 479-495.

Winters, S., Dubuc, C., & Higham, J. P. (2015). Perspectives: The looking time experimental paradigm in studies of animal visual perception and cognition. *Ethology*, *121*, 625-640.

Woods, R. J., & Wilcox, T. (2013). Posture support improves object individuation in infants. *Developmental Psychology, 49*, 1413-1424.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1-34.

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development, 22*, 145-160.

Woodward, A. L., Phillips, A., & Spelke, E. S. (1993). Infants’ expectations about the motion of animate versus inanimate objects. *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society* (pp. 1087-1091). Hillsdale, NJ: Erlbaum.

Woodward, A. L., Sommerville, J. A., & Guajardo, J. J. (2001). How infants make sense of intentional action. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds), *Intentions and Intentionality: Foundations of Social Cognition* (pp. 149-169). Cambridge, MA: MIT Press.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358,* 749-750.

Wynn, K. (2007). Some innate foundations of social and moral cognition. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The innate mind: Vol. 3 Foundations and the future* (pp. 330-347). Oxford: Oxford University Press.

Xu, F. (2007). Sortal concepts, object individuation, and language. *Trends in Cognitive Sciences, 11,* 400-406.

Xu, F., & Baker, A. (2005). Object individuation in 10-month-old infants using a simplified manual search method. *Journal of Cognition and Development, 6,* 307-323.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30,* 111-153.

Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology, 49,* 155-190.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences, USA, 105,* 5012-5015.

Zhang, F., & Emberson, L. L. (2020). Using pupillometry to investigate predictive processes in infancy. *Infancy, 25,* 758-780.

Zhang, Y, & Wang, S. (2019, September). *Violation to infant faulty knowledge induces object exploration by 7.5-month-olds in support events*. Paper presented at the Biennial Meeting of the Cognitive Development Society, Louisville, KY.

Ziv, T., & Sommerville, J. A. (2017). Developmental differences in infants' fairness expectations from 6 to 15 months of age. *Child Development, 88,* 1930-1951.

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology, 106,* 208-220.