Received: 17 February 2022 | Revised: 8 June 2022 | Accepted: 21 June 2022

DOI: 10.1002/icd.2359

COMMENTARY

WILEY

Great expectations: The construct validity of the violation-of-expectation method for studying infant cognition

Aimee E. Stahl[1] | Melissa M. Kibbe[2]

[1]Department of Psychology, The College of New Jersey, Ewing, New Jersey, USA

[2]Department of Psychological & Brain Sciences, Center for Systems Neuroscience, Boston University, Boston, Massachusetts, USA

Correspondence
Melissa M. Kibbe, Department of Psychological & Brain Sciences, Center for Systems Neuroscience, Boston University, Boston, MA, USA.
Email: kibbe@bu.edu

Aimee E. Stahl, Department of Psychology, The College of New Jersey, Ewing, New Jersey, USA.
Email: stahla@tcnj.edu

Abstract

The violation-of-expectation method has been used in thousands of studies examining the breadth and depth of preverbal infants' knowledge and cognitive capacities. In this commentary, we review evidence that supports the violation-of-expectation method as an important tool for infant research that has strong construct validity. We show that this method reveals infants' expectations, produces reliable results across study designs and stimuli, and is grounded in theories of learning. We support and encourage the use of this method to continue to unveil the scope and sophistication of infant cognition.

KEYWORDS
cognition, infancy, methodology, violation-of-expectation

The violation-of-expectation (VOE) method is an experimental technique used to assess the cognitive capacities of preverbal infants. The logic of the method involves showing infants events that either are consistent or inconsistent with an adult's expectations, and then comparing infants' behavioural or physiological responses to these different events. If infants differentially respond to the contrasting events, that is taken as evidence that infants had expectations about the outcome of the events. That is, infants detect a mismatch between what they *expected* and what they *observed* (otherwise known as prediction error). These expectations may have been present prior to the experimental session (e.g., Baillargeon et al., 1985; Gergely et al., 1995; McCrink & Wynn, 2004; Spelke et al., 1995; Téglás et al., 2007; Téglás et al., 2011; Wynn, 1992; Xu & Denison, 2009; Xu & Garcia, 2008), or acquired during the experimental session (e.g., Kirkham et al., 2002; Saffran et al., 1996; Stahl et al., 2014). The VOE method has been used in thousands of studies examining a range of domains of early cognition (for reviews, see, e.g., Baillargeon, 2008, Baillargeon et al., 2010, Baillargeon et al., 2012; Kibbe, 2015; Spelke & Kinzler, 2007; Saffran & Kirkham, 2018; Southgate, 2020). VOE methods are therefore thought to provide insights into the knowledge and cognitive capacities of infants who otherwise would not be able to verbally report those abilities.

In a recent commentary, Paulus (2022) suggested that infant researchers abandon the use of the VOE method. Paulus joins others (e.g., Cohen & Marks, 2002; Haith, 1998; Jackson & Sirois, 2009) who have raised concerns about

the construct validity of the VOE method, arguing that it is not possible to conclude that infants possess the kinds of knowledge attributed to them based on the results of VOE studies. In this commentary, we draw on several decades of converging evidence to argue instead that infants' behavioural and physiological responses in such studies are indeed reflective of infants' detection of prediction errors, and therefore of infants' cognitive capacities. We address some of the most common concerns leveraged against using the VOE method for infant research and show that the evidence supports the VOE method as a valid, reliable, and useful tool for characterizing early infant knowledge and learning.

1 | THE VOE METHOD YIELDS CONSISTENT RESULTS ACROSS DEPENDENT MEASURES

One of the main arguments leveraged against the VOE method is that it is not possible to interpret infants' nonverbal behaviours in these tasks as indicative of their underlying expectations. This critique is largely aimed at one of the most common nonverbal dependent measures used in VOE studies: infants' gaze duration following display events (i.e., "looking time"). The looking time measure capitalizes on the notion that preverbal infants will look longer at things that violate their expectations relative to things that accord with their expectations. For example, infants look longer when an object that moves behind an occluder appears to have passed through a solid barrier that was behind the occluder, relative to when the object is stopped by the barrier (e.g., Baillargeon, 1986; Baillargeon et al., 1985; Spelke et al., 1992). Based on infants' looking behaviour, it is assumed that infants possess expectations that objects should conform to the physical laws of solidity and spatiotemporal continuity. Some researchers have expressed scepticism that infants' longer looking times actually reflect their underlying expectations, or that infants could possess such expectations in the first place, and that therefore the large number of VOE studies that use looking time as their dependent measure are invalid or uninterpretable (e.g., Haith, 1998; Paulus, 2022; Tafreshi et al., 2014; see also Aslin, 2007).

Clues to the construct validity of VOE studies as a means of gauging infants' expectations come from the wide-ranging body of VOE studies that have used other dependent measures (both behavioural and neurophysiological), either instead of or in addition to looking time. In this section, we review this extensive converging evidence that infants' nonverbal behavioural and neurophysiological responses in VOE studies, including looking times, do indeed reflect their expectations.

Strong evidence for the construct validity of the VOE method comes from a newly emerging line of research that examines the functional utility of infants' behaviours following violations of expectation. When infants witness unexpected events, they are more likely to engage in targeted learning and exploration (e.g., Perez & Feigenson, 2022; Sim & Xu, 2017; Stahl & Feigenson, 2015; Stahl & Woods, in press; see Stahl & Feigenson, 2019 for review). Stahl and Feigenson (2015) presented infants with VOE events involving physical violations that typically elicit longer looking (e.g., a rolling car appeared to pass through a wall in its path), and found that infants more effectively learned novel information about objects that violated their expectations than objects that accorded with their expectations. They also showed that infants selectively explored objects that had behaved surprisingly over novel objects, and infants tested specific hypotheses about those surprising objects – infants were more likely to bang objects that appeared to violate principles of object solidity, but were more likely to drop objects that appeared to defy gravity. Multiple follow-up experiments ruled out perceptual novelty and increased arousal as alternative explanations for infants' surprise-induced learning and exploration. When an explanation for the surprising event is provided after the VOE event (e.g., there was a hole in the wall through which the car could pass), infants no longer preferred the surprising object, suggesting that infants' behaviours following VOE events were efforts to uncover an explanation for the unexpected event (Perez & Feigenson, 2022; see also Aguiar & Baillargeon, Aguiar & Baillargeon, 2002). Infants' ability to detect violations using the VOE method also is stable over time and predicts later measures of curiosity in early childhood (Perez & Feigenson, 2021). Taken together, these results confirm that infants are not simply responding passively to perceptual aspects of VOE events when they look longer at surprising outcomes, but rather are detecting violations of their expectations and treating such events as meaningful learning opportunities.

Other dependent measures also provide convincing evidence that the VOE method indeed taps into infants' expectations. For instance, infants experience changes to brain activity that reflect error detection in response to VOE events (e.g., Berger et al., 2006; Emberson et al., 2015; Köster et al., 2020; Kouider et al., 2015; Wilcox et al., 2005). As one example, Wynn (1992) showed infants an object that was covered by a screen, and a second object was subsequently added behind the screen. Infants looked longer when only one object was revealed behind the screen, relative to the expected two objects. Berger et al. (2006) showed infants the same events and found not only that infants looked longer at the violation event, but also that their event-related potentials reflected frontal activity associated with error detection. That is, infants' longer looking times indeed reflect a mismatch between what they expected and what they observed. Moreover, VOE studies have shown that infants process certain actions as goal-directed (e.g., Biro & Leslie, 2007; Luo & Baillargeon, 2005; Csibra, 2008; Woodward, 1998, 1999), and infants' brains similarly distinguish between actions that are goal-directed and those that are not (e.g., Nyström et al., 2011; Karthik et al., 2022; Southgate, Johnson, et al., 2010). These results suggest that infants' looking behaviours meaningfully reflect their cognitive experiences.

Studies that have examined infants' active behaviours also show convergence with infants' looking behaviours in VOE tasks. For example, not only do infants look longer at VOE events, but they also engage in more social referencing following such events (e.g., Dunn & Bremner, 2017; Walden et al., 2007). Infants have been shown to represent hidden objects via longer looking to VOE events (e.g., Baillargeon et al., 1985; Spelke et al., 1992; Wynn, 1992), but also via actively reaching and searching for them (e.g., Feigenson et al., 2002; Feigenson & Carey, 2003). Infants have demonstrated an understanding of others' intentions via longer looking in VOE studies (e.g., Behne et al., 2005; Gergely et al., 1995; Luo & Baillargeon, 2005; Woodward, 1998, 1999), which is mirrored in their imitation of others' goals (e.g., Carpenter, Akhtar, & Tomasello, 1998; Carpenter, Nagell, et al., 1998; Gergely et al., 2002; Hamlin et al., 2008; Meltzoff, 1995) and in their active interventions to correct others' predicted actions that would otherwise lead to errors based on false beliefs (e.g., Buttelmann et al., 2009; Knudsen & Liszkowski, 2012a). Finally, evidence from studies using anticipatory looking measures show that infants actively make predictions using knowledge acquired prior to or within the experimental session (e.g., Applin & Kibbe, 2019; Brandone et al., 2014; Cannon & Woodward, 2012; Fawcett & Liszkowski, 2012; Hunnius & Bekkering, 2010; Kim & Song, 2015; Kochukhova & Gredebäck, 2010; Romberg & Saffran, 2013; Ruffman et al., 2005; Saffran & Kirkham, 2018; Tummeltshammer & Kirkham, 2013). In a study that combined anticipatory looking and VOE looking time in a single experiment, Téglás and Bonatti (2016) found that infants anticipated the most probable motion path of an occluded object, and looked longer when the object followed a different path (see also Applin & Kibbe, 2019). These results suggest that infants are making active predictions, which could lead to behaviours such as longer looking when those predictions are violated.

Taken together, this convergent evidence corroborates the interpretation that infants differentially respond in VOE tasks because they detected something unexpected – that is, VOE methods measure infants' expectations, and therefore can yield insights into infants' cognitive capacities. Of course, we do not argue that violations of expectation are the *only* reason infants would differentially respond to an event (e.g., see Aslin, 2007; Bremner & Dunn, 2020; Mireault & Reddy, 2020). Instead, we note the wealth of converging evidence that validates the range of dependent measures commonly used in VOE studies, including looking time, active exploration and anticipatory measures, and neurophysiological measures. We join others in encouraging researchers to employ multiple dependent measures in VOE methods to best assess infant cognition (e.g., Bremner & Dunn, 2020; Byers-Heinlein et al., 2021; Havron, 2022; LoBue et al., 2020; Zettersten et al., 2022).

2 | THE VOE METHOD YIELDS CONSISTENT RESULTS ACROSS VARIATIONS IN TASKS

Another common argument against the use of the VOE method often cites alternative explanations for infants' differential behaviours in VOE events that do not require infants to have specific knowledge or cognitive capacities

(e.g., Cohen & Marks, 2002; Heyes, 2014; Paulus, 2022; Rivera et al., 1999; Schöner & Thelen, 2006). However, arguments along these lines often suffer from an aperture problem: they typically leverage these critiques against individual experiments or small sets of experiments, often positing a variety of different explanations for the results of individual studies, while failing to take into account the consistent patterns that emerge from larger sets of studies.

Specifically, some have argued that there are alternative low-level perceptual explanations for infants' behavioural or physiological responses to unexpected events (e.g., Bogartz et al., 1997; Haith, 1998; Heyes, 2014; Paulus, 2022; Schöner & Thelen, 2006). These arguments typically attribute infants' differential behaviours in VOE studies to perceptual properties of the display and/or perceptual novelty. Yet, the VOE method has yielded reliably consistent results across numerous studies within particular knowledge domains (e.g., physical reasoning) despite employing different experimental designs and stimuli (e.g., Baillargeon, 1986; Baillargeon, 1987; Baillargeon et al., 1985; Luo et al., 2003; Spelke et al., 1992; Spelke et al., 1995; Wang et al., 2004; Wilcox et al., 1996). Lower-level perceptual explanations of infants' longer looking in VOE tasks also are ruled out within the sets of VOE experiments presented in a typical journal article, which often employ the use of careful control conditions to rule out lower-level perceptual explanations (e.g., Fawcett & Liszkowski, 2012; Huntley-Fenner et al., 2002; Xu & Denison, 2009; Xu & Garcia, 2008; see also Munakata, 2000). For example, Kibbe and Leslie (2019) found that 6-month-old infants looked longer when an occluded object was revealed to have changed conceptual identity (either from a human-like doll face to a non-human ball, or vice versa). However, when infants were presented with a nearly identical experiment except with the doll face inverted (such that the lower-level perceptual differences between the stimuli were maintained while the conceptual distinctions between the objects were removed), infants did not look longer at the identity change outcome (see also Bonatti et al., 2002).

While it is possible that a unique perceptual explanation could be generated for each of these findings based on the specifics of those particular stimuli, these unique perceptual explanations ignore the emergent pattern of consistent behaviour elicited from these tasks. We do not argue that there are *no* VOE results that could potentially be attributed to lower-level perceptual explanations. Rather, we argue that carefully designed VOE studies, taken and interpreted together, can and do provide clear evidence for infant knowledge and cognitive capacities. We join others (e.g., Aslin, 2000; Kominsky et al., 2020; Scott & Baillargeon, 2014; Spelke, 1998; Spelke & Newport, 1998) who have argued that, when taken as a whole, the most parsimonious explanation of such studies is to attribute infants with the conceptual knowledge that aligns with their behaviours following violation events (e.g., that infants have expectations about how objects should behave).

The importance of interpreting the results of multiple VOE studies holistically is underscored by the individual VOE studies or sets of studies that have *failed* to replicate original findings. For example, Onishi and Baillargeon (2005) found that infants looked longer at events in which an agent acted unexpectedly given her false belief about the location of an object, suggesting potential evidence for a cognitive capacity for theory of mind in infancy. However, in the years following this finding, other studies have yielded mixed results, with some studies corroborating the notion that infants have false belief understanding in infancy using a variety of dependent measures (e.g., Buttelmann et al., 2009, 2015; Kovacs et al. Kovács et al., 2010, Knudsen & Liszkowski, 2012a, 2012b; Scott & Baillargeon, 2009; Song & Baillargeon, 2008; Southgate, Chevallier, et al., 2010; Southgate & Vernetti, 2014; Surian & Geraci, 2012; Surian et al., 2007; Träuble et al., 2010), and others obtaining null or opposite results (e.g., Crivello & Poulin-Dubois, 2018; Dorrenberg et al., Dörrenberg et al., 2018; Kampis et al., 2021; Phillips et al., 2015; Poulin-Dubois & Yott, 2018; Yott & Poulin-Dubois, 2016). Similar mixed results have been obtained in studies examining prosocial expectations in infants (see Margoni & Surian, 2018, for a meta-analysis).

There are many reasons that a study may fail to replicate. Some attempts at replicating VOE studies have made substantive changes to the original method that could potentially undermine the construct validity of the replication by, for example, using animated stimuli about which infants may not have strong a priori expectations (see Barr, Barr & Hayne, 1999; Barr, 2013; Diener et al., 2008; Ruysschaert et al., 2013), stimuli that were substantially different than the original studies in crucial ways (e.g., Oakes & Cohen, 1990; Ganglmayer et al., 2019; Paulus et al., 2011;

Sirois & Jackson, 2012; see Aslin, 2000 and Baillargeon et al., 2018 for critical reviews), and/or changing the method in ways that could lead to increased task demands (Keen, 2003 for discussion). And some studies may fail to replicate because the effect they measured is simply not real. As in all fields of experimental science, there may be a small subset of studies that are truly non-replicable, and the file-drawer problem may mask more failures to replicate than are currently known (Rosenthal, 1979). However, failures to replicate a small portion of VOE studies in a particular subdomain do not mean that *all* VOE studies are therefore invalid, or that the VOE method should never be used again. Just as in any science, we argue that infancy researchers must consider the evidence as a whole, including careful evaluation of the construct validity of replication attempts, and not base our theories on the outcome of a single study or paper (e.g., Maxwell et al., 2015). We encourage researchers to conduct and use replications, both small scale (individual replication attempts) and large scale (e.g., Byers-Heinlein et al., 2021; Visser et al., 2022), in building theories of infant cognition.

3 | THE VOE METHOD IS GROUNDED IN OUR BEST THEORIES OF LEARNING

The ability to detect violations of expectation is critical to navigating the world. Indeed, the idea that infants treat violations of expectation as opportunities to learn is grounded in our best theories of learning. Many dominant theories of learning are based on learners making predictions and subsequent revisions after observing evidence. Classic models of associative learning (e.g., Fletcher et al., 2001; Friston, 2005; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Schultz & Dickinson, 2000), Bayesian models of cognitive development (e.g., Perfors et al., 2011; Schulz, 2012; Tenenbaum et al., 2006), and predictive processing models of early learning (e.g., Andersen et al., 2022; Köster et al., 2020) all argue that prediction errors drive learning. While these theories differ in many respects, they all have at their cores the assumption that when there is a mismatch between what a learner expects and what they observe, they must update their behaviour or knowledge accordingly. That is, violations of expectations form the basis of learning (see Stahl & Feigenson, 2015, 2019).

4 | CONCLUSIONS

Decades of research support the construct validity of the VOE method for understanding the breadth and depth of infant cognition. The VOE method is a powerful tool that can be included in the infant researcher's toolkit, compatible with a variety of dependent measures (e.g., looking time, exploration, explanation-seeking, social referencing, physiological changes, active behaviours, and anticipatory looking), and compatible with our best theories of learning.

AUTHOR CONTRIBUTIONS

Aimee E. Stahl: Conceptualization (equal); writing - original draft preparation (equal); writing - review and editing (equal). **Melissa M. Kibbe:** Conceptualization (equal); writing - original draft preparation (equal); writing - review and editing (equal).

ACKNOWLEDGEMENTS

The authors wish to thank the members of the Boston University Developing Minds Lab for helpful discussion. This paper was supported in part by the Gitenstein-Hart Sabbatical Prize at The College of New Jersey to A.E.S.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/icd.2359.

ORCID

Aimee E. Stahl https://orcid.org/0000-0001-6585-1779

Melissa M. Kibbe https://orcid.org/0000-0002-9088-2523

REFERENCES

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

Andersen, M., Kiverstein, J., Miller, M., & Roepstorff, A. (2022). Play in predictive minds: A cognitive theory of play. Preprint https://psyarxiv.com/u86qy/

Applin, J. B., & Kibbe, M. M. (2019). Six-month-old infants predict Agents' goal-directed actions on occluded objects. *Infancy*, *24*(3), 392–410.

Aslin, R. N. (2000). Why take the cog out of infant cognition? *Infancy*, *1*(4), 463–470.

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

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. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, *3*(1), 2–13.

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., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

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

Baillargeon, R., Stavans, M., Wu, D., Gertner, Y., Setoh, P., Kittredge, A. K., & Bernard, A. (2012). Object individuation and physical reasoning in infancy: An integrative account. *Language Learning and Development*, *8*(1), 4–46.

Barr, R. (2013). Memory constraints on infant learning from picture books, television, and touchscreens. *Child Development Perspectives*, *7*(4), 205–210.

Barr, R., & Hayne, H. (1999). Developmental changes in imitation from television during infancy. *Child Development*, *70*(5), 1067–1081.

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology*, *41*, 328–337.

Berger, A., Tzur, G., & Posner, M. I. (2006). Infant brains detect arithmetic errors. *Proceedings of the National Academy of Sciences*, *103*, 12649–11653.

Biro, S., & Leslie, A. M. (2007). Infants' perception of goal-directed actions: Development through cue-based bootstrapping. *Developmental Science*, *10*(3), 379–398.

Bogartz, R. S., Shinskey, J. L., & Speaker, C. J. (1997). Interpreting infant looking: The event set× event set design. *Developmental Psychology*, *33*(3), 408–422.

Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology*, *44*(4), 388–426.

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*(1), 23–34.

Bremner, G., & Dunn, K. (2020). What's in a look? How can we best measure infants' response to incongruity? *Human Development*, *64*(2), 64–67.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342.

Buttelmann, F., Suhrke, J., & Buttelmann, D. (2015). What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology*, *131*, 94–103.

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

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, *15*(2), 292–298.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*, 315–330.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (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*, i–174.

Cohen, L. B., & Marks, K. S. (2002). How infants process addition and subtraction events. *Developmental Science*, *5*(2), 186–201.

Crivello, C., & Poulin-Dubois, D. (2018). Infants' false belief understanding: A non-replication of the helping task. *Cognitive Development*, *46*, 51–57.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*(2), 705–717.

Diener, M. L., Pierroutsakos, S. L., Troseth, G. L., & Roberts, A. (2008). Video versus reality: Infants' attention and affective responses to video and live presentations. *Media Psychology*, *11*(3), 418–441.

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, *46*, 12–30.

Dunn, K., & Bremner, J. G. (2017). Investigating looking and social looking measures as an index of infant violation of expectation. *Developmental Science*, *20*(6), e12452.

Emberson, L. L., Richards, J. E., & Aslin, R. N. (2015). Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months. *Proceedings of the National Academy of Sciences*, *112*, 9585–9590.

Fawcett, C., & Liszkowski, U. (2012). Infants anticipate others' social preferences. *Infant and Child Development*, *21*(3), 239–249.

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

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object-files versus analog magnitudes. *Psychological Science*, *13*, 150–156.

Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, R. A., Donovan, T., Papadakis, N., & Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature Neuroscience*, *4*, 1043–1048.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*, 815–836.

Ganglmayer, K., Attig, M., Daum, M. M., & Paulus, M. (2019). Infants' perception of goal-directed actions: A multi-lab replication reveals that infants anticipate paths and not goals. *Infant Behavior and Development*, *57*, 101340.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*(6873), 755.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193.

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

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*(4), 487–494.

Havron, N. (2022). Using multiple measures to improve reliability in infant. *Infant and Child Development*, e2336.

Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science*, *17*(5), 647–659.

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, *46*(2), 446–454.

Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small groups of discrete entities. *Cognition*, *85*, 203–221.

Jackson, I., & Sirois, S. (2009). Infant cognition: Going full factorial with pupil dilation. *Developmental Science*, *12*, 670–679.

Kampis, D., Karman, P., Csibra, G., Southgate, V., & Hernik, M. (2021). A two-lab direct replication attempt of Southgate, Senju and Csibra (2007). *Royal Society Open Science*, *8*(8), 210190.

Karthik, S., Parise, E., & Liszkowski, U. (2022). Mirroring communicative actions: Contextual modulation of mu rhythm desynchronization in response to the 'Back-of-hand' action in 9-month-old infants. *Developmental Neuropsychology*, *47*(3), 158–174.

Keen, R. (2003). Representation of objects and events: Why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science*, *12*(3), 79–83.

Kibbe, M. M. (2015). Varieties of visual working memory representation in infancy and beyond. *Current Directions in Psychological Science*, *24*(6), 433–439.

Kibbe, M. M., & Leslie, A. M. (2019). Conceptually rich, perceptually sparse: Object representations in 6-month-old infants' working memory. *Psychological Science*, *30*(3), 362–375.

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

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.

Knudsen, B., & Liszkowski, U. (2012a). Eighteen-and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science*, *15*(1), 113–122.

Knudsen, B., & Liszkowski, U. (2012b). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*, *17*(6), 672–691.

Kochukhova, O., & Gredebäck, G. (2010). Preverbal infants anticipate that food will be brought to the mouth: An eye tracking study of manual feeding and flying spoons. *Child Development*, *81*(6), 1729–1738.

Kominsky, J. F., Lucca, K., Thomas, A. J., Frank, M. C., & Hamlin, K. (2020). *Simplicity and validity in infant research*. https://psyarxiv.com/6j9p3/

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

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature Communications*, *6*, 1–7.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834.

LoBue, V., Reider, L. B., Kim, E., Burris, J. L., Oleas, D. S., Buss, K. A., … Field, A. P. (2020). The importance of using multiple outcome measures in infant research. *Infancy*, *25*(4), 420–437.

Luo, R., Baillargeon, R., Brueckner, L., & Munakata, Y. (2003). Reasoning about a hidden object after a delay: Evidence from robust representations in 5-month-old infants. *Cognition*, *88*, B23–B32.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, *16*(8), 601–608.

Margoni, F., & Surian, L. (2018). Infants' evaluation of prosocial and antisocial agents: A meta-analysis. *Developmental Psychology*, *54*(8), 1445–1455.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487–498.

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

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*, 838–850.

Mireault, G. C., & Reddy, V. (2020). Making sense of infants' differential responses to incongruity. *Human Development*, *64*(2), 55–63.

Munakata, Y. (2000). Challenges to the violation-of-expectation paradigm: Throwing the conceptual baby out with the perceptual processing bathwater? *Infancy*, *1*(4), 471–477.

Nyström, P., Ljunghammar, T., Rosander, K., & von Hofsten, C. (2011). Using mu rhythm desynchronization to measure mirror neuron activity in infants. *Developmental Science*, *14*(2), 327–335.

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, *5*(2), 193–207.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *8*, 255–258.

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

Paulus, M., Hunnius, S., van Wijngaarden, C., Vrins, S., van Rooij, I., & Bekkering, H. (2011). The role of frequency information and teleological reasoning in infants' and adults' action prediction. *Developmental Psychology*, *47*(4), 976–983.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*, 532–552.

Perez, J., & Feigenson, L. (2021). Stable individual differences in infants' responses to violations of intuitive physics. *Proceedings of the National Academy of Sciences*, *118*(27), 1–3.

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

Perfors, A., Tenenbaum, J. B., Griffith, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*, 302–321.

Phillips, J., Ong, D. C., Surtees, A. D., Xin, Y., Williams, S., Saxe, R., & Frank, M. C. (2015). A second look at automatic theory of mind: Reconsidering Kovács, Téglás, and Endress (2010). *Psychological Science*, *26*(9), 1353–1367.

Poulin-Dubois, D., & Yott, J. (2018). Probing the depth of infants' theory of mind: Disunity in performance across paradigms. *Developmental Science*, *21*(4), e12600.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Rivera, S. M., Wakeley, A., & Langer, J. (1999). The drawbridge phenomenon: Representational reasoning or perceptual preference? *Developmental Psychology*, *35*(2), 427–435.

Romberg, A. R., & Saffran, J. R. (2013). Expectancy learning from probabilistic input by infants. *Frontiers in Psychology*, *3*, 610.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.

Ruffman, T., Slade, L., & Redman, J. (2005). Young infants' expectations about hidden objects. *Cognition*, *97*(2), B35–B43.
Ruysschaert, L., Warreyn, P., Wiersema, J. R., Metin, B., & Roeyers, H. (2013). Neural mirroring during the observation of live and video actions in infants. *Clinical Neurophysiology*, *124*(9), 1765–1770.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203.
Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, *113*(2), 273–299.
Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*, *23*, 473–500.
Schulz, L. E. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*, 382–389.
Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, *80*(4), 1172–1196.
Scott, R. M., & Baillargeon, R. (2014). How fresh a look? A reply to Heyes. *Developmental Science*, *17*(5), 660–664.
Sim, Z., & Xu, F. (2017). Infants preferentially approach and explore the unexpected. *British Journal of Developmental Psychology*, *35*, 596–608.
Sirois, S., & Jackson, I. R. (2012). Pupil dilation and object permanence in infants. *Infancy*, *17*(1), 61–78.
Song, H. J., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, *44*(6), 1789–1795.
Southgate, V. (2020). Are infants altercentric? The other and the self in early social cognition. *Psychological Review*, *127*(4), 505–523.
Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*(6), 907–912.
Southgate, V., Johnson, M. H., Karoui, I. E., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological Science*, *21*(3), 355–359.
Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130*(1), 1–10.
Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, *21*, 181–200.
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., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, *13*, 113–142.
Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.
Spelke, E. S., & Newport, E. L. (1998). Nativism, empiricism, and the development of knowledge. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Volume 1: Theoretical models of human development* (5th ed., pp. 275–340). John Wiley & Sons, Inc..
Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*, 91–94.
Stahl, A. E., & Feigenson, L. (2019). Violations of core knowledge shape early learning. *Topics in Cognitive Science*, *11*, 136–153.
Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development*, *85*, 1821–1826.
Stahl, A. E., & Woods, L. (in press). Infants preferentially learn from surprising teachers. *Infancy*.
Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580–586.
Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, *30*(1), 30–44.
Tafreshi, D., Thompson, J. J., & Racine, T. P. (2014). An analysis of the conceptual foundations of the infant preferential looking paradigm. *Human Development*, *57*(4), 222–240.
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*, *104*, 19156019159.
Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*, 1054–1058.
Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *7*, 309–318.
Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs? *Infancy*, *15*(4), 434–444.
Tummeltshammer, K. S., & Kirkham, N. Z. (2013). Learning to look: Probabilistic variation and noise guide infants' eye movements. *Developmental Science*, *16*(5), 760–771.

Visser, I., Bergmann, C., Byers-Heinlein, K., Dal Ben, R., Duch, W., Forbes, S., Franchin, L., Frank, M., Geraci, A., Hamlin, J. K., Kaldy, Z., & Zettersten, M. (2022). Improving the generalizability of infant psychological research: The ManyBabies model. *Behavioral and Brain Sciences*, *45*, e35.

Walden, T., Kim, G., McCoy, C., & Karrass, J. (2007). Do you believe in magic? Infants' social looking during violations of expectations. *Developmental Science*, *10*(5), 654–663.

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.

Wilcox, T., Bortfeld, H., Woods, R., Wruck, E., & Boas, D. A. (2005). Using near-infrared spectroscopy to assess neural activation during object processing in infants. *Journal of Biomedical Optics*, *10*, 11010-1–11010-9.

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

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*(2), 145–160.

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

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, *112*, 97–104.

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

Yott, J., & Poulin-Dubois, D. (2016). Are infants' theory-of-mind abilities well integrated? Implicit understanding of intentions, desires, and beliefs. *Journal of Cognition and Development*, *17*(5), 683–698.

Zettersten, M., Pomper, R., & Saffran, J. (2022). Valid points and looks: Reliability and validity go hand-in-hand when improving infant methods. *Infant and Child Development*, *e2326*.

How to cite this article: 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. https://doi.org/10.1002/icd.2359