

Online measures of looking and learning in infancy

Alexis S. Smith-Flores^{1,2}  | Jasmin Perez²  |
Michelle H. Zhang²  | Lisa Feigenson² 

¹Department of Psychology, University of California, San Diego, La Jolla, California, USA

²Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, Maryland, USA

Correspondence

Alexis S. Smith-Flores, Department of Psychology, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093, United States
Email: alexis-smith@ucsd.edu

Abstract

Infants in laboratory settings look longer at events that violate their expectations, learn better about objects that behave unexpectedly, and match utterances to the objects that likely elicited them. The paradigms revealing these behaviors have become cornerstones of research on pre-verbal cognition. However, little is known about whether these canonical behaviors are observed outside laboratory settings. Here, we describe a series of online protocols that replicate classic laboratory findings, detailing our methods throughout. In Experiment 1a, 15-month-old infants ($N = 24$) looked longer at an online support event culminating in an Unexpected outcome (i.e., appearing to defy gravity) than an Expected outcome. Infants did not, however, show the same success with an online solidity event. In Experiment 1b, 15-month-old infants ($N = 24$) showed surprise-induced learning following online events—they were better able to learn a novel object's label when the object had behaved unexpectedly compared to when it behaved expectedly. Finally, in Experiment 2, 16-month-old infants ($N = 20$) who heard a valenced utterance (“Yum!”) showed preferential looking to the object most likely to have generated that utterance. Together, these results suggest that, with some adjustments, online testing is a feasible and promising approach for infant cognition research.

1 | INTRODUCTION

For several decades, infants' looking responses in laboratory experiments have provided a window on early human cognition and perception (see Aslin, 2007; Spelke & Kinzler, 2007). Measuring what infants look at, when, and for how long, has enabled researchers to make inferences about early emerging abilities and expectations, long before infants can engage in tasks that require language or instruction. Experiments measuring looking time employ a range of paradigms, including violation of expectation (VOE), habituation, anticipatory looking, object-sound matching, and preferential looking, and have yielded rich information and provided fodder for lively debate regarding what infants know about objects, quantities, language, categories, and other people.

An outstanding question for developmental psychology concerns the generality of these responses: can infant looking times only be reliably captured in the laboratory, or can they also be obtained using virtual methods, in infants' everyday environments? This question is especially pressing at present, as the COVID-19 global pandemic has shuttered research laboratories around the world. In-person infant testing has largely ceased, for now. Must infant looking-time research also be put on hold?

Even before the first news of COVID-19, efforts were being made to develop purely online platforms for conducting research with infants and young children. In one especially broadly scaled effort, Scott and Schulz (2017) created an automated online platform called *Lookit*, designed to allow researchers to test participants asynchronously. To participate in a *Lookit* experiment, parents log on to the website at their convenience, without the presence or assistance of an experimenter, and follow instructions to initiate presentation of the test stimuli. The webcam in parents' computers transmits video of infants' looking or children's responses, which can be coded later by the research team. This automation feature necessitates that experiments on *Lookit* be self-explanatory in their setup and robust across a range of parameters (e.g., participants' screen size, viewing distance and angle, lighting conditions, audio quality). Scott and colleagues reported an initial comparison of data collected using *Lookit*, relative to in-lab experiments, in three experiments. These involved VOE with 14-month-old infants, preferential looking with 2-year-olds, and word learning with 4-year-olds (Scott et al., 2017). The first two are most relevant to the present work, as they measured looking time. In the VOE experiment, adapted from Téglás et al. (2007), infants watched four objects bouncing in a container—three were identical and one differed in shape and color. The container was covered, and then one of the objects “randomly” emerged from an opening in the bottom. In the original study, 12-month-old infants looked longer when the unique object emerged, compared to one of the majority objects. Scott et al.'s (2017) replication also found longer looking to this improbable outcome compared to the probable one; however, the difference was considerably smaller than the in-lab study and did not reach statistical significance (even with more than double the sample size of the original study). In the preferential looking experiment, based on Yuan and Fisher (2009), 2-year-olds' syntactic bootstrapping abilities were tested by presenting them with a video of an action involving a single person and a video of an action involving two people, and asking them to find a novel verb (e.g., “Find blinking!”), that earlier had been used in either a transitive or an intransitive sentence. Children in Scott and colleagues' online replication showed similar looking patterns to children in Yuan and Fisher's in-person study. Overall, these results suggest that positive looking-time results can be obtained with toddlers, but leave open whether looking-time studies with infants can be run successfully online.

Outside of the *Lookit* platform, other researchers have used Amazon Mechanical Turk (MTurk) for online infant recruitment and testing. Tran et al. (2017) showed infants various clips from children's television programs and measured their looking using automated online testing without

an experimenter, similar to *Lookit*. Rather than seeking to replicate any particular in-lab finding, Tran and colleagues sought to determine simply whether parents would volunteer their infants to participate in a study on MTurk, whether infants would be engaged by videos on a home computer, and whether the online recordings of infants' looking would be of sufficient quality to support later coding. They found that they were able to capture infants' looking time and that most infants in their sample were engaged by the stimuli (and by some of the videos in particular). Tran et al. also noted that although online data collection proceeded more quickly than in-person data collection (with 147 5- to 8-month-olds tested over a two-week period), 61.2% of their sample (90 infants) had to be excluded because of technical issues (e.g., low video quality, poor infant placement in the video). The high exclusion rate was similar to that in Scott et al.'s *Lookit* study (2017), in which data from 58.4% of infants in the VOE experiment (157 of 269 infants) and 57.4% of toddlers in the preferential looking experiment (189 of 329 toddlers) were excluded from the final analysis.

No other published studies that we are aware of have tested infants' looking time online. However, there are other online studies with older children that used measures other than looking. Rhodes and colleagues (Rhodes et al., 2020) developed an unmoderated platform similar to *Lookit*, in that parents and children participate without guidance from a live experimenter. They report a replication of a study by Rhodes et al. (2012), which asked whether young school-age children (4–8 years) exhibit an increase in essentialist beliefs after hearing statements using generic language. In their large, multi-national sample, they observed a similar pattern to that in the original in-person study. Lo et al. (2021) also created a browser-based platform for running unmoderated infant and child studies. They report a replication of a study by Mani and Huettig (2012), which found that some 2-year old children can use constrained verbs (i.e., “eat”) to predict which word will likely follow (i.e., “cake”).

Finally, some experimenters have recently demonstrated success at testing children online, but with a live experimenter rather than an automated protocol. Sheskin and Keil's (2018) platform, *The Child Lab*, tested children in an online version of a false belief task (Baron-Cohen et al., 1985), a resource distribution task assessing children's intuitions about fairness (Sheskin et al., 2016), and a plinko machine task in which children tried to predict object trajectories (Hood, 1995). Overall, the 50 children tested (ranging from 4 to 12 years old) performed well in the fairness and plinko tasks (with at least 46 children answering correctly in each), but were less successful in the false belief task, with only 33 children correctly identifying where a character with a false belief would search for a hidden object (for other developmental theory-of-mind data collected online, see Smith-Flores & Feigenson, 2021). These results suggest that online testing with a live experimenter is feasible, although it apparently does not always produce results that are fully equivalent to in-lab testing.

The recent advances in testing reviewed above suggest that it may be possible to capture infants' looking patterns online, in their own homes, even in the midst of a pandemic. However, an outright success at using online methods to replicate a classic infant looking-time result has yet to be observed. The “live experimenter” methods of Sheskin and Keil (2018) lend credence to the idea that laboratory tasks can be recreated, at least to some degree, off-campus, retaining the social interaction with the experimenter that may help children understand the task or feel engaged, and helping parents adjust their cameras or computers to optimize data transmission.

Our goal in the present work was to develop a protocol that could feasibly merge these two ideas: testing infants online, using a live video-chat method. Here, we tested infants in three classic infant looking-time paradigms, online with a live experimenter. The paradigms we used were chosen because ongoing work in our laboratory was using these methods at the time of the transition to online testing. Experiment 1a used a violation of expectation (VOE) paradigm based on previous

work by Needham and Baillargeon (1993) and Spelke et al. (1992). We showed infants videos of two events: one involving object support and the other involving object solidity. Each infant saw one event culminate in an Expected outcome (i.e., an object moved while remaining fully supported, or a moving object was stopped by a wall in its path), and one event culminate in an Unexpected outcome (i.e., an object moved completely off its supporting surface and appeared to hover in mid-air, or an object appeared to pass straight through a wall in its path). Based on previous work, we Expected infants to look longer at the Unexpected than the Expected outcomes. In Experiment 1a, we measured infants' looking following object events with Expected versus Unexpected outcomes, as in laboratory experiments by Stahl and Feigenson (2015). Following the object event, infants from Experiment 1b were taught a novel label for the object that had just accorded with or defied expectations. Then, their learning of the novel label was measured. Based on previous work, we expected infants to learn better following the Unexpected events. Finally, in Experiment 2, we tested infants in a preferential looking-time paradigm similar to that of Wu et al. (2017), which measured infants' looking following valenced utterances. Infants heard an experimenter say either "Yum!" or "Oh!" (a neutral utterance) while they saw an appetizing stimulus (like cake) and a positive but non-edible stimulus (a toy). Based on previous work, we expected infants to increase their looking to the appetizing stimulus when hearing "Yum!" but not "Oh!".

2 | GENERAL METHOD

2.1 | Recruitment

Participants were recruited from a database of children whose parents had previously expressed interest in laboratory research opportunities via local recruitment (e.g., farmers' markets, pre-schools, museums) and online advertising. Parents of infants between 13 and 19 months of age received an email inviting them to participate in a 20-minute online research study with their child. After parents had scheduled a virtual appointment, they received an email with a meeting link, a separate link to an online consent form, and a link to an optional demographics form. Parents were told that they could wait to sign the consent form until their appointment if they had any questions. All forms were hosted and completed through Qualtrics. Parents were emailed a \$5 Amazon gift card after their visit.

2.2 | Consent

Parents signed an electronic consent form approved by the university Internal Review Board (IRB); at the end of the form, parents typed their full legal name to give consent for their child to participate and for the testing session to be recorded. The form also included a PDF version of the consent form that parents could download if they wished. The last question on the form asked about the type of computer parents would use during the visit in order to collect data about screen size constraints.

When the online appointment began, the experimenter introduced themselves and briefly explained the study's aim and methods. If parents had not already signed the consent form, the experimenter asked whether the parent had questions about the form or the study, and then, when parents were ready, resent the consent form link. If parents had already signed the consent form, the experimenter asked whether they had any further questions, reiterated that the study would be video-recorded, and reminded parents that the study could be stopped at any time.

Online study recordings were saved directly to the lab's secure video server, which was only accessible through the university's virtual private network (VPN).

3 | ONLINE PROTOCOL

3.1 | Video-chat setup

Given the importance of capturing fine-grained eye movements, and previous reports of high numbers of participant exclusions in other online studies (Scott et al., 2017; Tran et al., 2017), our primary concern about online testing was capturing high quality data. To maximize data quality, we designed an online testing protocol that allowed the experimenter to troubleshoot in real time, just as they would if the study were run in a campus laboratory. The live testing procedure gave parents the opportunity to ask questions and allowed the experimenter to help parents modify their setup. For example, the experimenter sometimes asked parents to move infants closer to or farther from the screen, adjust the lighting, or move infants from a highchair to the parent's lap if they become fussy. These modifications helped the experimenter capture higher quality data. Additionally, the live format of the experiment allowed the experimenter to pause the testing session if infants became upset. Depending on when the pause occurred (e.g., during a pre-trial attention getter or between test blocks), the study could resume after infants had calmed down. Because of the variety of screen sizes across households, a calibration video was included at the beginning of each testing session to capture footage of infants' looks to the center, left, and right sides of their screen.

3.2 | Zoom meeting setup

We chose Zoom as our video-chat platform because many parents were anticipated to already be familiar with it. We pre-tested the quality of our stimulus videos among research team members via Zoom screen-sharing, in order to determine whether the videos were of sufficient quality for online testing.¹ Prior to each Zoom testing session, we set up the recording output so that it would capture both the video stimuli and an image of the participant. Our settings enabled the footage of participants to be positioned next to the stimulus footage, and for the entire recording to be saved directly to the lab's secure video server rather than on a personal computer. For details on the settings we used, see the Supporting Information (and for additional resources for testing children online using live experimenter-mediated approaches, see the open access materials created by the Social Learning Lab at Stanford University (2020)).

Once on the Zoom call, parents were provided more details about the study. The experimenter then prompted them to position infants in a highchair, or on their lap if a highchair was not

¹We highly recommend pre-testing stimuli online prior to piloting the actual study. If, after testing video stimuli on Zoom via screen-sharing, the videos are judged to be of poor quality (e.g., the video lags, images are distorted, sound is not synced), researchers may wish to consider slides.com (Slides) as an alternative platform for hosting stimuli. Slides supports online slide show presentation, with video stimuli uploaded to the website, inserted onto a series of slides, and displayed live. The benefit of this format is that the presented videos are of higher quality, with reduced lags between video and audio. In addition, although we do not discuss infant controlled paradigms here, slides can be used to present stimuli using an infant controlled procedure rather than fixed looking windows. For studies that use slides for stimulus presentation, Zoom can be modified to record testing sessions (see Supporting Information for instruction manual on Zoom setup).

available or suitable for the position of the parent's computer. Parents were further instructed to remain silent throughout the study and to avoid interacting with infants (e.g., not to comment, point at the screen, or otherwise redirect infants' attention). Once infants were positioned close enough to the screen to capture their gaze, but far away enough so that infants could not reach the keyboard or computer mouse, the experimenter began recording and initiated the stimulus video.

3.3 | Exclusion criteria

Participants were excluded for similar reasons as would apply during laboratory testing (e.g., fussiness, parental interference, experimenter error). We also excluded infants who looked less than two seconds during any measurement period (indicating a general lack of interest in the stimuli) or did not look within the first five seconds of the event looking window (see Experiments 1–3, below). Two additional exclusion criteria were added as a result of online testing; these concerned video quality and screen size. First, infants' data were excluded if their video recording had a noticeable lag between the stimuli's audio and video (as it became difficult to know when coding for a particular measurement period should begin), or if the quality of the video recordings was too poor to clearly see where infants were looking. Additionally, we excluded any participant whose screen size measured less than 10 inches or greater than 24 inches diagonally. Screens that were too small precluded us from being able to accurately code where infants were looking; screens that were too large required infants to turn their heads to see some of the presented stimuli, thereby also making coding difficult.

3.4 | Experiments 1A and 1B

Experiment 1 was divided into two phases. Experiment 1a sought to determine whether infants would show the classic pattern of heightened looking in response to a VOE, in this case involving the behavior of physical objects. We tested 15-month-old infants using an adapted version of the VOE paradigm in the first experiment by Stahl and Feigenson (2015). Infants' looking was compared for an event in which expectations about object support were or were not violated, and an event in which expectations about object solidity were or were not violated. In Experiment 1b, we asked whether surprise-induced learning is also observed in online testing. Participants were the same infants from Experiment 1a. After infants had the opportunity to observe the Unexpected or Expected Outcome in Experiment 1a, we taught them a novel label for the object that had participated in the event; we then tested infants' learning of these novel labels.

4 | EXPERIMENT 1A: VIOLATION OF EXPECTATION IN SOLIDITY AND SUPPORT EVENTS

4.1 | Method

4.1.1 | Participants

Twenty-four 13- to 19-month-old infants participated (M_{age} : 15.68 months, range: 13.15–19.00 months, SD: 1.75 months; 13 girls). We tested infants across a relatively wide range of ages in

order to begin to assess the generality of our online testing results. Nineteen infants sat in a high-chair and five sat on their caregiver's lap. Data from 25 additional infants were excluded for screen size constraints (4), fussiness (7), interference from a household member (5), low quality video (3), or looking time during event windows being less than two seconds (6). Infants excluded from Experiment 1a were also excluded from Experiment 1b. One child was identified by their caregiver as Black or African American, 19 as White, and four as racially mixed. Two caregivers indicated their child was Hispanic/Latinx. The modal screen size across participants was 13 inches (range: 11–21.5 inches). The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Homewood Institutional Review Board at the Johns Hopkins University.

4.1.2 | Design and stimuli

Each infant participated in a total of two trials, one involving a solidity event and one involving a support event. One of these events ended in an Expected outcome and the other in a Unexpected outcome. Whether the Unexpected outcome was associated with a support or a solidity event was counterbalanced across infants, as was whether infants saw the support event first or second. This yielded four possible combinations of event outcome pairings and orders.

In the support event, infants saw two novel objects, both easily visually distinguishable from one another: a round green dog toy and a long orange dog toy. The green toy measured 8 cm in diameter, and the orange toy measured 15 cm long \times 7 cm wide. Which of these served as the Target Object and which served as the Distractor Object was counterbalanced across participants.

In the solidity event, infants saw two novel objects with hidden wheels (so that they could be rolled). One had a yellow wicker top with a pink felt belt around its lower half (16.5 \times 13 \times 14.5 cm), and the other was made of gray felt and had a pink hair curler positioned vertically on top, with a blue pipe cleaner wrapped around its perimeter (17 \times 10 \times 16 cm). For half the infants, the yellow wicker object was the Target Object and the gray object was the Distractor, and for the other half, this pairing was reversed.

Infants saw the digitally recorded events unfold on a black puppet stage. A black curtain was lowered to cover the stage between trials.

4.1.3 | Procedure

Calibration

The study began with a brief calibration event in which an image of a colorful beach ball bounced in the middle of the screen; then, a fish bounced on the left side of the screen, and finally, a different fish bounced on the right side of the screen (in the locations where the test stimuli would later appear). Each bounce was accompanied by a *Boing!* sound to draw infants' attention. This calibration window provided coders with an image of what it looked like when each infant was looking in the key locations of the screen.

Support event

During the support event, infants first saw a black curtain raised to reveal the novel green and orange objects stationary on a puppet stage, with a black occluding screen resting between

them. The experimenter pointed to both objects simultaneously, one with each hand, and said, “Look at these!” She then grasped an object in each hand, picked them up, and placed them down again on the stage. The video zoomed in on each object so infants could see it up close. Then, the curtain was lowered; this marked the end of the 20-s object familiarization event. Infants next saw the curtain raised again to reveal each object positioned atop a gray block on either side of the stage. Whether the Target Object was on the left or right was counterbalanced across participants. The experimenter said, “Look at this!” as she reached down and touched the Target Object with her index finger (Figure 1, Panel 1). The experimenter then either pushed the Target Object to the edge of the block, with the object remaining fully supported as it moved (Expected Outcome; Figure 1, Panel 2A) or pushed the Target Object completely off the block, leaving it suspended in mid-air (Unexpected Outcome; Figure 1, Panel 2B). Infants’ looking at the outcome after the experimenter removed her finger was measured for 15 s.

Solidity event

The solidity Event began with a short familiarization trial, designed to introduce infants to the stimulus objects and the ramp. Infants first saw two novel objects on an empty stage. As in the support event, the camera zoomed in on each object, to show them to infants up close. Then, the curtain was lowered; this marked the end of the 20-s object familiarization event. Next, infants saw the curtain raised to reveal a black ramp (120 cm long) slanting downward from the left side of the puppet stage. At the opposite end of the stage was a gray wall. Infants saw the experimenter reach down and place a black occluding screen on stage, partially concealing the path between the end of the ramp and the gray wall. The curtain was lowered and then immediately raised to reveal the two objects onstage: the Target Object atop the ramp, and the other object resting stationary in the left corner of the stage. The experimenter said, “Look at this!”, reached down and pointed to the Target Object, then said, “Watch this!” and pushed it so that it rolled down the ramp and behind the occluding screen. She then said, “Look at this!” as she removed the screen, revealing the Target Object resting against the gray wall on the far-right side of the stage. After five seconds, the experimenter said, “Down we go!” and lowered the curtain.

Next, in the critical event trial, the curtain was again raised to reveal the ramp, the black occluding screen covering part of the trajectory from the end of the ramp to the right end of the stage, and both objects in the same positions they had been in during familiarization (the Target Object atop the ramp and the other object on the left side of the stage floor). In addition, a highly salient red wall could be seen protruding 13 cm above the top of the occluding screen (Figure 2, Panel 1). The experimenter drew attention to the red wall by running her finger down its side while saying, “Look at this!” Next, she said, “Watch this!” and pushed the Target Object so that it rolled down the ramp and passed behind the occluding screen. After two seconds, the experimenter said, “Look at this!” as she removed the occluding screen to reveal either the Target Object resting against the red wall—as if the red wall had stopped it (Expected Outcome; Figure 2, Panel 2A)—or against the far gray wall—as if it had passed through the red wall (Unexpected Outcome; Figure 2, Panel 2B).

Infants’ total looking at the computer screen during the 15 s immediately following the end of each event was coded offline by two experienced observers. Twenty percent of the video footage was coded by both observers; their intercoder reliability was 95.6%.

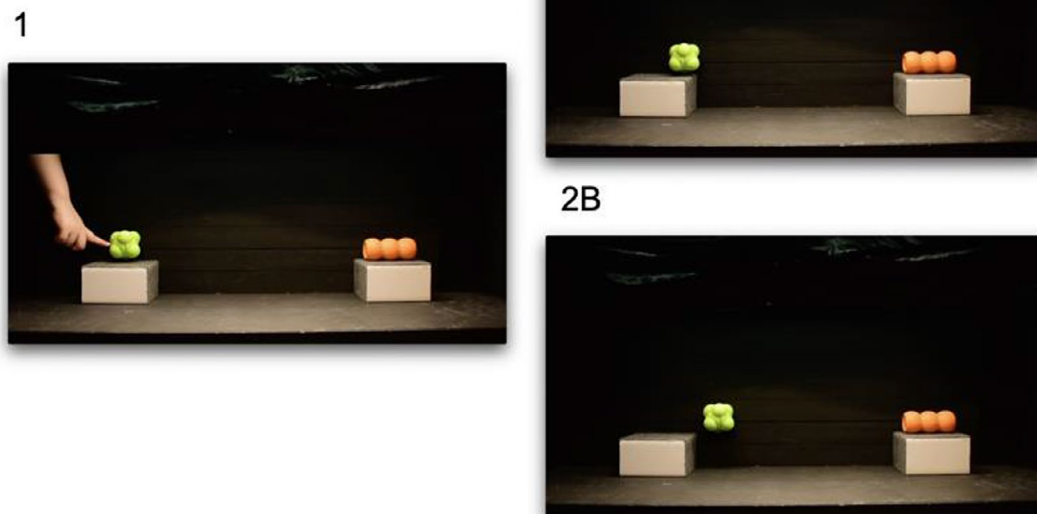


FIGURE 1 Schematic of Experiment 1a support event. *Note.* The experimenter drew infants' attention to one of the novel objects (Panel 1). She then either pushed it to the edge of the block, leaving it supported (Expected Outcome, Panel 2A) or pushed it off the block completely, leaving it apparently floating in mid-air (Unexpected Outcome, Panel 2B)

4.2 | Results

Infants looked at the Unexpected Outcome of a Support event for an average of 12.5 s (SD = 2.38 s), and at the Expected Outcome of a Support event for average of 9.86 s (SD = 2.7 s). Infants looked at the Unexpected Outcome of a Solidity event for an average of 9.41 s (SD = 3.48 s), and at the Expected Outcome of a Solidity event for an average of 10.61 s (SD = 3.8 s).

All analyses were planned ahead of data collection and were modeled after previous experiments (see Perez & Feigenson, 2020; Stahl & Feigenson, 2015). We examined infants' looking times using a repeated measures ANOVA with Outcome Type as a within-subject factor (Unexpected vs. Expected), and Outcome Event Pairing (a Solidity event with an Unexpected Outcome paired with a Support event with an Expected Outcome, or a Solidity event with an Expected Outcome paired with a Support event with an Unexpected Outcome) and Outcome Order (Unexpected Outcome or Expected Outcome first) as between-subject factors. This revealed no main effect of Outcome Type, $F(1, 20) = 1.44$, $p = .24$, $\eta^2 = .07$, nor of Outcome Event Pairing, $F(1, 20) = 2.58$, $p = .124$, $\eta^2 = .11$, or Outcome Order, $F(1, 20) = 0.45$, $p = .51$, $\eta^2 = .02$. However, we observed a marginally significant interaction between Outcome Type and Outcome Event Pairing, $F(1, 20) = 3.35$, $p = .08$, $\eta^2 = .14$: the difference in infants' looking at Unexpected versus Expected Outcomes was marginally greater for infants who saw the Support event culminate in an Unexpected Outcome, compared to infants who saw the Solidity event culminate in an

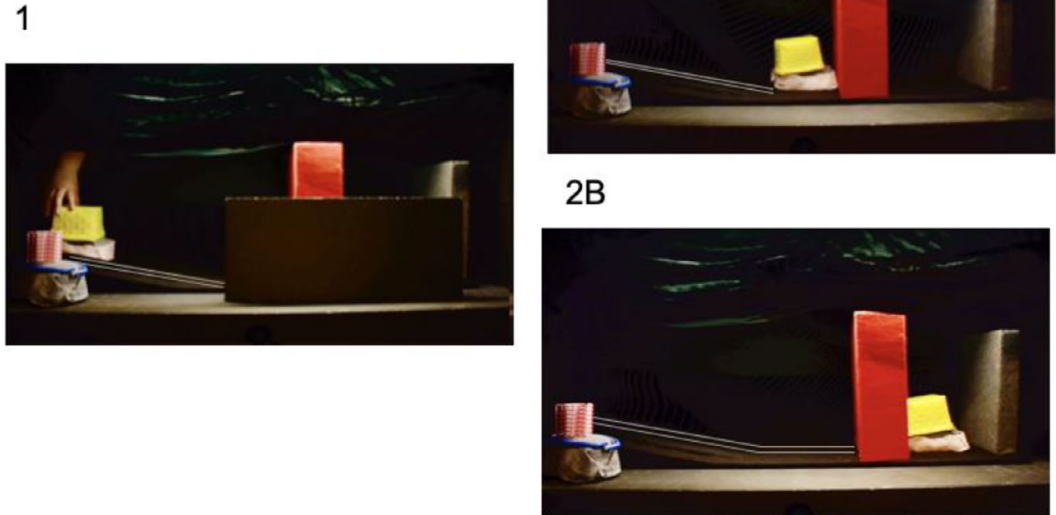


FIGURE 2 Schematic of Experiment 1a solidity event. *Note.* Infants saw the Target Object atop the ramp and the other object sitting nearby (Panel 1). After the Target Object was pushed down the ramp, the black occluding screen was raised to reveal the Target Object on the near side of the red wall (Expected outcome, Panel 2A) or on the far side of the red wall (Unexpected Outcome, Panel 2B)

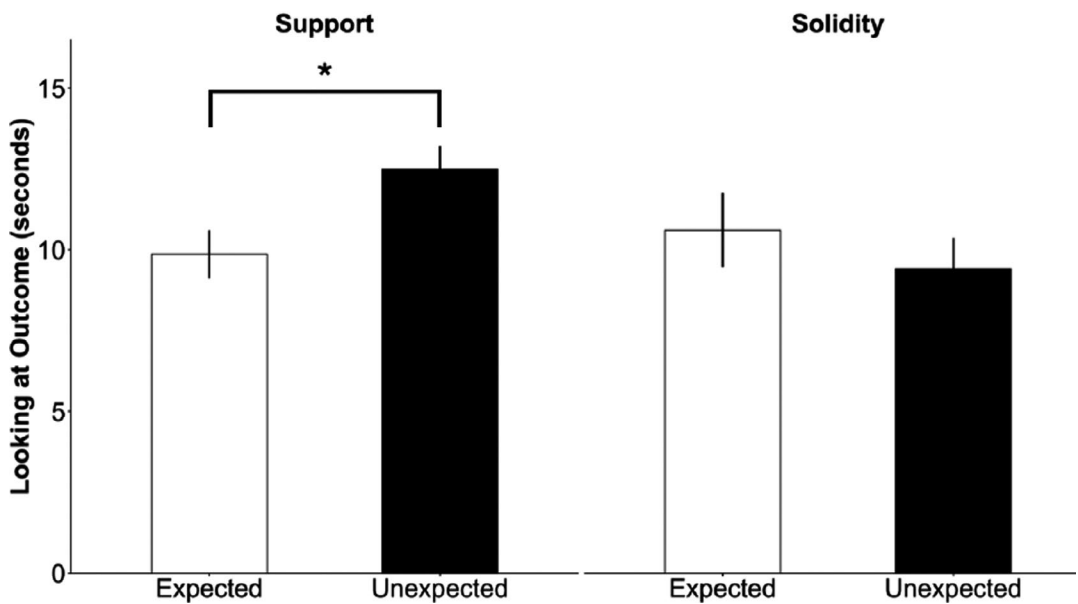


FIGURE 3 Infants' looking times in Experiment 1a. *Note.* Error bars reflect SEM. * $p = .02$

Unexpected Outcome (Figure 3). No other interactions were significant. We further examined infants' looking using planned t tests. We found that infants looked longer at the Unexpected than the Expected outcome of the support event, $t(22) = 2.52$, $p = .02$, 95% CI [0.46, 4.81]; in contrast, they did not show this elevated looking to the Unexpected outcome of the solidity event, $t(22) = -0.81$, $p = .43$, [-4.28, 1.88].

The results of Experiment 1a offer evidence that VOE studies can be successfully run online, at least for some event types. Infants looked longer when an object appeared to hover in mid-air, compared to when the object remained fully supported. This success was observed despite the fact that infants received just a single opportunity to observe the support event, rather than the multiple trials that are typical of laboratory studies (Needham & Baillargeon, 1993). However, contrary to our predictions, we did not observe a successful VOE response to an online solidity event—infants failed to look longer when a moving object appeared to have passed through a wall, compared to when it was stopped by the wall. We speculate on possible reasons for this failure in the General Discussion.

5 | EXPERIMENT 1B: LEARNING FOLLOWING UNEXPECTED VS. EXPECTED OUTCOMES OF SUPPORT AND SOLIDITY EVENTS

In addition to looking longer at violations-of-expectation, infants and young children also show heightened learning following these Unexpected events, at least in laboratory settings (Stahl & Feigenson, 2015, 2017). For example, 11-month-old infants saw an object appear to pass through a solid wall or saw an object appear to violate spatiotemporal continuity by being hidden in Location A but being retrieved at Location B. Immediately following these Unexpected outcomes, infants were taught new information about the object in the event (that it made a novel sound). Infants learned this new information better, compared to when the same events culminated in Expected outcomes (i.e., the object was stopped by the wall, or the object hidden in Location A was retrieved in Location A). In Experiment 1b, we asked whether learning following Unexpected object events is also observed outside of the laboratory in studies conducted online.

5.1 | Method

5.1.1 | Participants

The same infants from Experiment 1a participated in Experiment 1b.

5.1.2 | Procedure

Infants previously saw two events: one support event and one solidity event (one culminating in an Unexpected Outcome and the other in an Expected Outcome). For each of these, after infants' looking at the outcome had been measured for 15 s, the stage curtain was lowered. Then, after 5 s, it was raised to reveal the Target Object (i.e., the object that had participated in the preceding support or solidity event) and the Distractor Object (i.e., the object that had been seen during the preceding event, but was an “innocent bystander” rather than the main focus of the



FIGURE 4 Schematic of Experiment 1b support event. *Note.* Infants saw the Target and Distractor Objects on either side of the stage for 5 s, during which baseline looking to each object was measured (Panel 1). The experimenter then pointed to and labeled the Target Object (Panel 2). Finally, infants were prompted to look at the Target Object using its novel label, and looking to each object was measured (Panel 3)

event) stationary on the left and right sides of the stage (spatial position counterbalanced across infants; Figure 4, Panel 1). Infants' looking was measured for 5 s—this provided a baseline index of infants' visual interest in each object. After this, the experimenter reached down, pointed to the Target Object, and said either, “That's a gaffa! Yeah, a gaffa!” (following the Support event; Figure 4, Panel 2), or “That's a diffy! Yeah, a diffy!” (following the Solidity event), then lowered the curtain.

Finally, infants' learning of the novel label was measured. The curtain was raised to reveal the Target and Distractor objects in the same spatial positions in which they had just been seen. The experimenter, who was not visible to infants, then said, “Where's the [gaffa/ diffy]? Look at the [gaffa/diffy]!” (Figure 4, Panel 3). Infants' looking was measured for 10 s, starting from the onset of the first utterance of the Target Object label. After 10 s, the curtain was lowered briefly and raised again to reveal the objects once more, now having switched locations on the stage. Again, infants were prompted to look at the Target Object by the offstage experimenter, as in the previous trial. This resulted in two learning test trials per event type (two trials measuring learning following an Unexpected outcome and two trials measuring learning following an Expected outcome).

Twenty percent of the video footage was coded by two experienced observers; intercoder reliability was 93.4%. Infants' raw looking times were then converted to learning scores (as in Stahl & Feigenson, 2015). First, we calculated the proportion of time infants spent looking at the Target Object during the baseline measurement period (i.e., the 5-s silent interval after both objects were revealed on stage, but before the experimenter labeled the Target), by dividing each infant's total looking at the Target Object by their total looking at the Target Object plus their total looking at the Distractor Object. Then, we calculated the proportion of time infants spent looking at the Target Object during the two test trials (i.e., the 10-s intervals following the onset of the Target Object's label), again by dividing each infant's total looking at the Target Object by their total looking at the Target Object plus their total looking at the Distractor Object. Finally, we subtracted the baseline proportion from the average proportion across the two test trials. A resulting positive learning score would indicate an increased tendency to correctly look at the Target Object when hearing it labeled, over and above any baseline preference to look at that object.

5.2 | Results

Infants' learning score following the Unexpected Outcome of a Support event was 0.12 ($SD = 0.28$), and their learning score following the Expected Outcome of a Support event was -0.22

(SD = 0.22). Infants' learning score following the Unexpected Outcome of a Solidity event was 0.05 (SD = 0.29), and their learning score following the Expected Outcome of a Solidity event was -0.01 (SD = 0.27).

We examined infants' learning scores using a repeated measures ANOVA, with Outcome Type (Unexpected vs. Expected) and Trial (first vs. second learning trial) as within subjects factors, and Outcome Event Pairing (a Solidity event with an Unexpected Outcome paired with a Support event with an Expected Outcome, or a Solidity event with an Expected Outcome paired with a Support event with an Unexpected Outcome) and Outcome Order (Unexpected Outcome or Expected Outcome first) as between-subject factors. This analysis revealed a significant main effect of Outcome Type, $F(1, 20) = 7.97, p = .01, \eta^2 = .29$; infants learned better following an Unexpected Outcome ($M = 0.09, SD = 0.28$) than an Expected Outcome ($M = -0.12, SD = 0.26$). We also observed a marginally significant main effect of Event Outcome Pairing, $F(1, 20) = 3.20, p = .089, \eta^2 = .14$; the difference in infants' learning following Unexpected versus Expected Outcomes was marginally greater for infants who had seen the Support event culminate in an Unexpected Outcome, compared to infants who saw the Solidity event culminate in an Unexpected Outcome. There were no other main effects or interactions.

As in Experiment 1a, we conducted planned comparisons to more closely examine infants' performance across the two event types. We found that infants learned the novel label better following a violation of object support, compared to following a nearly identical event that did not violate object support, $t(22) = 3.352, p = .003, 95\% \text{ CI } [0.13, 0.56]$ (Figure 5). In contrast, the difference in infants' learning following a violation of object solidity versus an event that did not violate object solidity was not significant $t(22) = -0.537, p = .594, [-0.17, 0.30]$ (Figure 5).

Thus, the results of Experiment 1b parallel those of Experiment 1a. Most critically, we found that we were able to replicate the effect of learning following object violations using online methods: infants learned significantly better following an Unexpected outcome of an object support

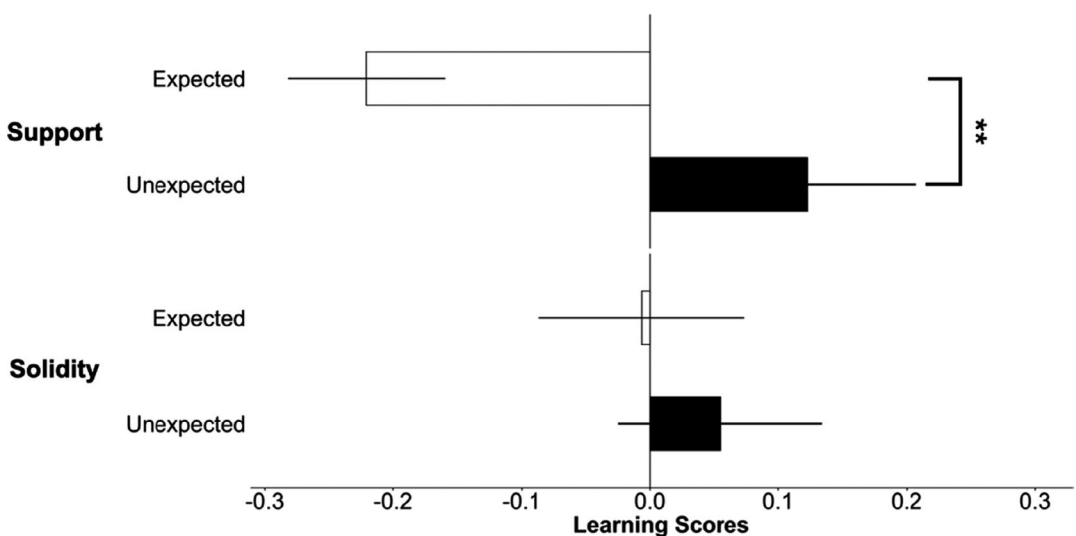


FIGURE 5 Infants' learning scores in Experiment 1b. *Note.* Positive learning scores indicate longer looking at the correct Target Object following object labeling. Error bars reflect SEM. $**p = .003$

event, compared to following an Expected outcome. Although there was no significant Outcome Type \times Outcome Event Pairing interaction in the ANOVA, simple t tests suggested that infants did not learn better following an Unexpected outcome of an object solidity event when tested online.

6 | EXPERIMENT 2: PREFERENTIAL LOOKING

A third classic paradigm in infant cognition research is the paired preference procedure. In this paradigm, infants see two contrasting stimuli while hearing auditory input, and their relative looking at the two images is measured. Experiments using this paradigm find that, for example, infants hearing someone utter a particular vowel sound look longer at faces whose lip shape matches rather than mis-matches that vowel sound (Patterson & Werker, 2003), infants look longer at a correct than an incorrect referent when hearing an object label (e.g., “Find the *shoe*”; “Do you see the *shoe*?”; Golinkoff et al., 1987; Reznick & Goldfield, 1992; Schafer & Plunkett, 1998), and infants look longer at an image of multiple novel objects than an image of a single novel object when hearing a sentence marking plurality (“Look, there are some blickets!” vs. “Look, there is a blicket!”; Kouider et al., 2006). Although Scott et al. (2017) demonstrated successful looking in an online paired preference task by 24- to 36-month-old children, it is not yet known whether preferential looking can be adapted to online testing with infants, who are too young to receive verbal task instructions.

In Experiment 2, we asked whether infants tested online could use a person's emotional utterance to determine which of two possible objects was its most likely elicitor. The method was based on a previous in-lab study by Wu et al. (2017), in which infants saw images of two objects that, for adults, were expected to elicit contrasting positive emotional responses (e.g., a dessert and a light up toy—one of which might be perceived as delicious, and the other as exciting or fun). Infants heard an emotion-relevant utterance (e.g., “Yum!”), and their looking was measured. Wu et al. (2017) found that 18- to 23-month-olds looked at the predicted image when they heard a given utterance in a laboratory setting (e.g., infants looked at the dessert after they heard “Yum!”). In our experiment, 16-month-old infants saw an image of a dessert on one side of the screen and an image of a stuffed animal toy on the other, then heard a female speaker either say “Yum!” or the neutral “Oh!”. We hypothesized that infants would look longer at the dessert when they heard “Yum!” but would have no looking preference between the two objects when they heard “Oh!”.

6.1 | Method

6.1.1 | Participants

Twenty 14- to 18-month-old infants participated (M: 16.46 months, range: 14.04–17.95 months, SD: 1.04 months; 15 girls). Nine infants sat in a highchair and 11 sat in a caregiver's lap. Data from one additional infant were excluded due to low video quality. One infant was identified by their parent as Asian, 18 as White, and one as being of mixed race; one infant was identified as Hispanic/Latinx. The modal screen size across participants was 15 inches (range: 13–20 inches). The study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before

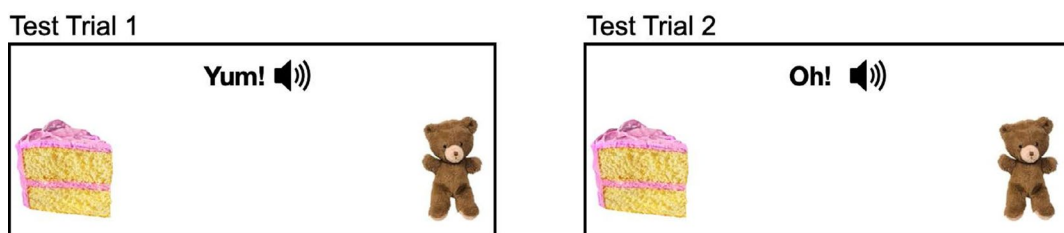


FIGURE 6 Schematic of Experiment 2

any assessment or data collection. All procedures involving human subjects in this study were approved by the Homewood Institutional Review Board at the Johns Hopkins University.

6.1.2 | Procedure

Calibration proceeded as in Experiment 1a—infants saw a series of animated stimuli appear in the center, left, and right of the screen, accompanied by attention-getting sounds. This served to engage infants with the screen and allowed the coders to see what it looked like when each infant was looking in the locations where the upcoming test stimuli would appear. Following calibration, each infant participated in two test blocks; each block was comprised of one baseline trial and two test trials. On one block infants saw a piece of pink-frosted cake on one side of the screen and a brown teddy bear on the other; on the other block, infants saw a brown cupcake on one side and a green toy turtle on the other (Figure 6). If infants saw the dessert on the left for the first block, they saw it on the right for the second block; order of side-object pairing was counterbalanced across participants.

Baseline trial

In each baseline event, the two objects appeared simultaneously and a recorded female voice said, “Look at these!” The objects remained visible for eight seconds, during which infants’ looking time was recorded. Then, the two objects disappeared, and a beach ball appeared and bounced twice in the middle of the screen, to redirect infants’ attention back to the center of the screen before the test trials began.

Test trials

On each test trial, infants saw the same two objects from the preceding baseline trial, in the same locations. After one second, they heard a recorded female voice say either “Yum!” or “Oh!” (Figure 6). Infants’ looking was recorded for the subsequent 10 s while the objects remained visible onscreen. After 10 s, the two objects disappeared and the beach ball appeared in the center of the screen and bounced twice again, to redirect infants’ attention. Then, infants were presented with the second test trial, which was identical to the first except for the utterance. If infants had heard “Yum!” on the first test trial, then they heard “Oh!” on the second, and vice versa.

Infants’ looking to each object was coded offline by two experienced observers. Twenty percent of the video footage was coded by both observers; intercoder reliability was 95.5%. Infants’ raw looking times were then converted to preference scores: we divided the time each infant spent looking at the dessert on each test trial by their total looking to either object during the test trial. We then subtracted from this the time infants spent looking at the dessert during the

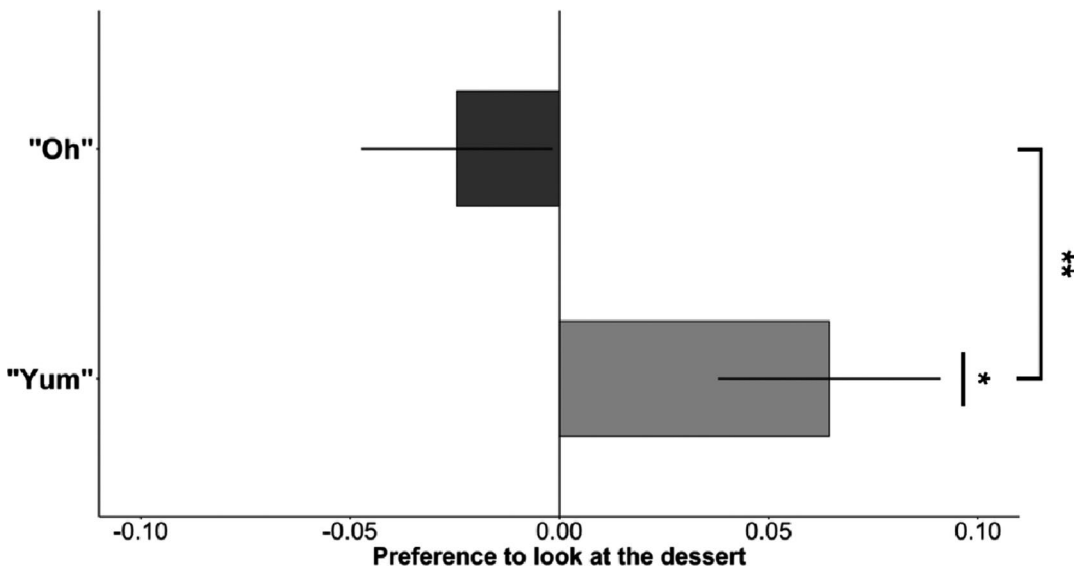


FIGURE 7 Infants' average preference to look to the dessert image after hearing "Yum!" or "Oh!" in Experiment 2. Note. Error bars reflect SEM. $*p = .024$, $**p = .008$

baseline trial divided by their total looking to either object during the baseline trial. This yielded a measure of increased preference during the test trial. A positive preference score would indicate an increased preference to look at the dessert following the utterance, relative to infants' preference to look at the dessert during baseline.

6.2 | Results

The proportion of time infants looked at the dessert item averaged 0.06 ($SD = 0.19$) after hearing "Yum!", and averaged -0.02 ($SD = 0.10$) after hearing "Oh!".

We analyzed infants' preference scores using a repeated measures ANOVA with Utterance ("Yum!" vs. "Oh!") and Trial (first vs. second instance of utterance) as within-subject factors, and Block Order (cake slice and teddy bear first vs. cupcake and turtle first) as a between-subjects factor. This revealed a significant main effect of Utterance, $F(1, 18) = 8.74$, $p = .008$, $\eta^2 = .33$, (Figure 7), with infants showing a greater preference for the dessert after hearing "Yum!" than after hearing "Oh!". As hypothesized, infants also looked longer at the dessert after hearing "Yum!" than would be expected by chance, $t(19) = 2.46$, $p = .024$, 95% CI [0.01, 0.12], whereas when they heard "Oh!" they looked equally at the two stimuli, $t(19) = -1.05$, $p = .31$, 95% CI $[-0.07, 0.02]$. We also observed a main effect of Block Order, $F(1, 18) = 5.37$, $p = .03$, $\eta^2 = .23$; infants whose first test block presented the cupcake and the turtle looked longer at the dessert across both blocks ($M = .07$, $SD = 0.16$), compared to infants whose first block presented the cake and the teddy bear ($M = -0.02$, $SD = 0.16$). No other main effects or interactions were observed.

Infants' longer looking at the dessert after hearing "Yum!" demonstrates that preferential looking can be successfully captured via online testing. Moreover, this evidence corroborates prior work, here demonstrating that younger 16-month-old infants have expectations about the kinds of objects that elicit certain responses. Critically, infants' at-chance looking when hearing

the neutral utterance “*Oh!*” confirms that infants did not simply have an overall preference to look at cakes and cupcakes, relative to non-food items.

7 | GENERAL DISCUSSION

In three experiments, we asked whether three different infant looking-time paradigms could be adapted to an online testing format using experimenter-mediated methods. In Experiment 1a, we found that infants looked longer at an event in which the principle of object support was violated, compared to an event that did not violate object support. However, the same infants showed no evidence of detecting a violation of object solidity—a result that contrasts with in-lab studies (Perez & Feigenson, 2020, 2021; Spelke et al., 1992). In Experiment 1b, we found that infants’ word learning was enhanced following a violation of object support, compared to following a nearly identical event that did not culminate in a violation. Finally, in Experiment 2, we found that infants looked longer at images that matched an emotional utterance (i.e., looked longer at a dessert than a toy when hearing “*Yum!*”). Together, these results suggest that infant looking-time paradigms can be successfully adapted for online testing.

Despite the successful findings reported above, our results also suggest that online testing differs in some important ways from in-lab testing. The failure of infants in Experiment 1a to show increased looking at a solidity violation implicates some important potential differences between our video-recorded stimuli and the live events that we and others have shown infants in the laboratory. We suspect that one difference between online and in-lab testing has to do with the visual experience of the test stimuli. Due to the screen size limitations of remote testing, future studies interested in capturing infants’ responses to object trajectories (including continuity violations, instances of the tunnel-effect, launching events) may need to think carefully about the visual angle of the stimuli, which can become distorted when seen on a screen. For example, the most common participant screen size in our experiments ranged from 13 to 15 inches. In this small space, the size ratio of the objects in the foreground of the scene (e.g., the target object) to the objects in the background (e.g., the ramp and walls in our solidity study) differed significantly from that in our laboratory setup. The distance traversed by the object in the solidity events of our Experiments 1 and 2 was much smaller than that in the laboratory, in terms of change in visual angle. This more minimal evidence of the object’s trajectory may have been partly responsible for the null results we observed in the solidity condition of our experiments. In comparison, in our support events, the object’s trajectory was less crucial to perceiving the event; the object’s lateral movement through space was not required to convey the sense of support versus lack of support from below. More work is clearly needed to closely examine such differences in testing formats. In addition, future work will ideally be able to provide quantitative comparisons of infants’ performance in the very same tasks conducted online versus in the laboratory. For example, it will be useful to compare effect sizes obtained in successful online studies versus their in-lab counterparts (e.g., in order to determine appropriate sample sizes for future work). Unfortunately, we were not able to conduct such an effect size comparison here, as our previous in-lab studies examining infants’ violation of expectation and learning responses tested infants of a different age than those tested here. However, as researchers begin to experiment more with online testing, we envision that a database of such comparisons will prove an invaluable resource for the field.

In addition, future online infant studies may wish to explore the use of infant controlled looking, as opposed to the fixed trial durations that we used, and that were also used by Scott et al.

(2017). Infant controlled looking allows trial lengths to vary based on individual infants' behavior; this may provide a more sensitive measure. Our method of using Zoom with a live experimenter could be easily adapted for this—for example, researchers can use slides.com or similar platforms to control stimulus presentation dynamically in real time (see Supporting Information for instructions).

It is unclear whether experimenter-mediated studies result in lower exclusion rates than asynchronous methods. Differences in the length and/ or stimuli of particular studies may play a larger role in determining attrition than the online versus in-person nature of the study, although this remains a question that will require more data to answer. In addition, the causes of participant attrition in online studies, and the impact of this attrition on the resulting data, merit further examination. As noted by Slaughter and Suddendorf (2007), the thresholds used to exclude participants from laboratory studies often are not well described in published work and likely differ across research groups. This problem may be exacerbated by the addition of online studies, because of the additional variability in testing conditions. For example, experimenters who are not in the room with infants may be more reliant on parents' reactions to infants' fussing when deciding whether to end the testing session early, and as such may be less able to apply the same standards equally across all infants in the test sample. Future work that makes comparisons between in-lab and online data collection should consider exclusion criteria as another key factor contributing to the data.

Still, our overall conclusion is that, based on our experience testing infants over the past year, online infant experiments are a viable way to make continued research progress when in-person laboratory testing is not possible. Furthermore, online testing need not be a replacement for in-lab testing; it can function in parallel, as an additional means to reach families who lack the means to travel to a campus laboratory, or whose schedules make laboratory visits inconvenient. Researchers who wish to recruit special populations, or to test children from racial, ethnic, or socioeconomic backgrounds that differ from the population geographically proximal to their laboratory also may wish to explore this testing format. Of course, whether online testing diversifies an experiment sample depends on the individual circumstances of the laboratory, and on the population they would otherwise have tested. It is also important to note that participants from families with access to home computers and reliable Internet, and with an interest in volunteering for experiments, are more likely to be White and from middle or upper-middle socioeconomic backgrounds (Lourenco & Tasimi, 2020). As such, the question of how research findings generalize outside the context of the experimental sample persists for online experiments, just as it long has for those conducted in the laboratory. However, recent online efforts to recruit children from beyond the local communities adjacent to university laboratories may help with this problem, at least to some degree. The relatively new website ChildrenHelpingScience.com hosts links to different online studies from laboratories around the world; caregivers from across the globe can sign up to have their children participate in online studies. Although this does not solve the problem of recruiting samples of children across the entire swath of socioeconomic backgrounds, online testing sites with broad reach may move our field closer to that goal.

Finally, here we focused on infant looking-time procedures and their feasibility for online testing, but we are aware that the right way to interpret the looking-time patterns observed in such studies remains the subject of healthy debate (Aslin, 2007; Haith, 1998; Spelke, 1998). Questions as to how to best understand infants' looking in any given study (e.g., Does it reflect knowledge? Expectations? Detection of perceptual novelty?) remain just as potent, and just as open, for online work as for work conducted in the laboratory. Our goal here was not to provide evidence on how such findings should best be interpreted; rather, we hope that introducing new means of

adding more empirical evidence to the discussion will serve to advance this critical dialogue in positive ways.


In summary, we see this as an exciting time for infant research. Our field is poised to embrace ambitious endeavors at scales that were unthinkable in the earlier decades of cognitive development research. Multi-lab replication efforts like ManyBabies (Byers-Heinlein et al., 2020, 2021), other types of large-scale collaborative networks (Sheskin et al., 2020), and online testing like that pioneered by *Lookit* (Scott et al., 2017; Scott & Schulz, 2017) and extended in the present work all have the potential to broaden the scope and impact of research on the developing mind. We hope our work can contribute to that effort.

ACKNOWLEDGMENTS


The authors declare no conflicts of interest with regard to the funding source for this study or otherwise.

ORCID

Alexis S. Smith-Flores  <https://orcid.org/0000-0002-6755-4986>

Jasmin Perez  <https://orcid.org/0000-0003-1543-5755>

Michelle H. Zhang  <https://orcid.org/0000-0001-9186-6712>

Lisa Feigenson  <https://orcid.org/0000-0002-9015-7065>

REFERENCES

- Aslin, R. N. (2007). What's in a look? *Developmental Science*, 10(1), 48–53. <https://doi.org/10.1111/j.1467-7687.2007.00563.x>
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37–46. [https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)
- Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., Kominsky, J. F., Kosie, J. E., Lew-Williams, C., Liu, L., Mastroberardino, M., Singh, L., Waddell, C. P. G., Zettersten, M., & Soderstrom, M. (2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology/Psychologie Canadienne*, 61(4), 349–363. <https://doi.org/10.1037/cap0000216>
- Byers-Heinlein, K., Tsui, A. S. M., Bergmann, C., Black, A. K., Brown, A., Carbajal, M. J., Durrant, S., Fennell, C. T., Fiévet, A.-C., Frank, M. C., Gampe, A., Gervain, J., Gonzalez-Gomez, N., Hamlin, J. K., Havron, N., Hernik, M., Kerr, S., Killam, H., Klassen, K., ... Wermelinger, S. (2021). A multilab study of bilingual infants: Exploring the preference for infant-directed speech. *Advances in Methods and Practices in Psychological Science*, 4(1), 251524592097462. <https://doi.org/10.1177/2515245920974622>
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(1), 23–45. <https://doi.org/10.1017/S030500090001271X>
- Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development*, 21(2), 167–179. [https://doi.org/10.1016/S0163-6383\(98\)90001-7](https://doi.org/10.1016/S0163-6383(98)90001-7)
- Hood, B. M. (1995). Gravity rules for 2-to 4-year olds? *Cognitive Development*, 10(4), 577–598. [https://doi.org/10.1016/0885-2014\(95\)90027-6](https://doi.org/10.1016/0885-2014(95)90027-6)
- Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, 2(1), 1–25. https://doi.org/10.1207/s15473341l1d0201_1
- Lo, C. H., Mani, N., Kartushina, N., Mayor, J., & Hermes, J. (2021). e-Babylab: An Open-source Browser-based Tool for Unmoderated Online Developmental Studies. *PsyArXiv*. <https://doi.org/10.31234/osf.io/u73sy>
- Lourenco, S. F., & Tasimi, A. (2020). No participant left behind: Conducting science during COVID-19. *Trends in Cognitive Sciences*, 24(8), 583–584. <https://doi.org/10.1016/j.tics.2020.05.003>
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843. <https://doi.org/10.1037/a0029284>

- Needham, A., & Baillargeon, R. (1993). Intuitions about object support in 4.5-month-old infants. *Cognition*, 47(2), 121–148. [https://doi.org/10.1016/0010-0277\(93\)90002-D](https://doi.org/10.1016/0010-0277(93)90002-D)
- Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6(2), 191–196. <https://doi.org/10.1111/1467-7687.00271>
- Perez, J., & Feigenson, L. (2020). *Violations of expectation trigger infants to search for explanations*. PsyArXiv. <https://doi.org/10.31234/osf.io/eahjd>
- 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), e2103805118. <https://doi.org/10.1073/pnas.2103805118>
- Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology*, 28(3), 406. <https://doi.org/10.1037/0012-1649.28.3.406>
- Rhodes, M., Leslie, S. J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, 109(34), 13526–13531. <https://doi.org/10.1073/pnas.1208951109>
- Rhodes, M., Rizzo, M. T., Foster-Hanson, E., Moty, K., Leshin, R. A., Wang, M., Benitez, J., & Ocampo, J. D. (2020). Advancing developmental science via unmoderated remote research with children. *Journal of Cognition and Development*, 21(4), 477–493. <https://doi.org/10.1080/15248372.2020.1797751>
- Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*, 69(2), 309–320. <https://doi.org/10.2307/1132166>
- Scott, K., Chu, J., & Schulz, L. (2017). Lookit (Part 2): Assessing the viability of online developmental research, results from three case studies. *Open Mind*, 1(1), 15–29. https://doi.org/10.1162/OPMI_a_00001
- Scott, K., & Schulz, L. (2017). Lookit (Part 1): A new online platform for developmental research. *Open Mind*, 1(1), 4–14. https://doi.org/10.1162/OPMI_a_00002
- Sheskin, M., & Keil, F. C. (2018). *TheChildLab.com: A video chat platform for developmental research*. PsyArXiv. <https://doi.org/10.31234/osf.io/rn7w5>
- Sheskin, M., Nadal, A., Croom, A., Mayer, T., Nissel, J., & Bloom, P. (2016). Some equalities are more equal than others: Quality equality emerges later than numerical equality. *Child Development*, 87(5), 1520–1528. <https://doi.org/10.1111/cdev.12544>
- Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online developmental science to foster innovation access, and impact. *Trends in Cognitive Sciences*, 24(9), 675–678. <https://doi.org/10.1016/j.tics.2020.06.004>
- Slaughter, V., & Suddendorf, T. (2007). Participant loss due to “fussiness” in infant visual paradigms: A review of the last 20 years. *Infant Behavior and Development*, 30, 505–514. <https://doi.org/10.1016/j.infbeh.2006.12.006>
- Smith-Flores, A., & Feigenson, L. (2021). Preschoolers represent others' false beliefs about emotions. *Cognitive Development*, 59, 101081. <https://doi.org/10.1016/j.cogdev.2021.101081>
- Social Learning Lab. (2020). Online test: Startup guide and materials. *Zenodo*. <https://doi.org/10.5281/zenodo.3762737>
- Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, 21(2), 181–200. [https://doi.org/10.1016/S0163-6383\(98\)90002-9](https://doi.org/10.1016/S0163-6383(98)90002-9)
- Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 99(4), 605. <https://doi.org/10.1037/0033-295X.99.4.605>
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 89–96. <https://doi.org/10.1111/j.1467-7687.2007.00569.x>
- Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, 348(6230), 91–94. <https://doi.org/10.1126/science.aaa3799>
- Stahl, A. E., & Feigenson, L. (2017). Expectancy violations promote learning in young children. *Cognition*, 163, 1–14. <https://doi.org/10.1016/j.cognition.2017.02.008>
- 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(48), 19156–19159.
- Tran, M., Cabral, L., Patel, R., & Cusack, R. (2017). Online recruitment and testing of infants with Mechanical Turk. *Journal of Experimental Child Psychology*, 156, 168–178. <https://doi.org/10.1016/j.jecp.2016.12.003>

- Wu, Y., Muentener, P., & Schulz, L. E. (2017). One-to four-year-olds connect diverse positive emotional vocalizations to their probable causes. *Proceedings of the National Academy of Sciences*, 114(45), 11896–11901. <https://doi.org/10.1073/pnas.1707715114>
- Yuan, S., & Fisher, C. (2009). “Really? She blicked the baby?” Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, 20(5), 619–626. <https://doi.org/10.1111/j.1467-9280.2009.02341.x>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher’s website.

How to cite this article: Smith-Flores, A. S., Perez, J., Zhang, M. H., & Feigenson, L. (2021). Online measures of looking and learning in infancy. *Infancy*, 00, 1–21. <https://doi.org/10.1111/infa.12435>