Developmental Psychology

ISSN: 0012-1649

2021, Vol. 57, No. 12, 2150–2164
https://doi.org/10.1037/dev0001253

Testimony Bias Lingers Across Development Under Uncertainty

Rista C. Plate[1], Kristin Shutts[2, 3], Aaron Cochrane[4], C. Shawn Green[2], and Seth D. Pollak[2, 3]

[1] Department of Psychology, University of Pennsylvania
[2] Department of Psychology, University of Wisconsin–Madison
[3] Waisman Center, University of Wisconsin–Madison
[4] Faculty of Psychology and Educational Sciences, University of Geneva

Children have a powerful ability to track probabilistic information, but there are also situations in which young learners simply follow what another person says or does at the cost of obtaining rewards. This latter phenomenon, sometimes termed bias to trust in testimony, has primarily been studied in children preschool-age and younger, presumably because reasoning capacities improve with age. Less attention has been paid to situations in which testimony bias lingers—one possibility is that children revert to a testimony bias under conditions of uncertainty. Here, participants (4 to 9 years old) searched for rewards and received testimony that varied in reliability. We find support for testimony bias beyond preschool-age, particularly for uncertain testimony. Children were sensitive to trial-by-trial uncertainty (Experiment 1: *N* = 102, 59 boys, 43 girls; the sample included nine Hispanic/Latinx, 93 non-Hispanic/Latinx participants, of whom six were Black/African American, seven were Asian American, eight were multiracial, 77 were White, and four indicated "other" or did not respond), and with *uncertainty* defined as a one-time, unexpected change in the testimony (Experiment 3: *N* = 129; 68 boys, 61 girls; the sample included 12 Hispanic/Latinx, 117 non-Hispanic/Latinx [10 Black/African American, four Asian American, nine multiracial, 103 White, and three "other"]). However, the impact of the testimony bias decreased with age. These effects were specific to the testimony coming from another person as opposed to resulting from a computer glitch (Experiment 2: *N* = 89, 52 boys, 37 girls; five Hispanic/Latinx, 80 non-Hispanic/Latinx, of whom one was Black/African American, three were Asian American, 15 were multiracial, 66 were White, and four did not report race). Taken together, these experiments provide evidence of a disproportionate influence of testimony, even in children with more advanced reasoning skills.

Keywords: testimony, uncertainty, social learning, development

Supplemental materials: https://doi.org/10.1037/dev0001253.supp

Children are powerful learners. Whereas adults can make inferences and predictions based on coherent and structured knowledge about the world that is formulated over time, early learners must begin formulating their knowledge of the world based on uncertain information. But what patterns of evidence do children rely upon to learn most effectively about the world? Children are not solitary learners: They attend to and follow other people, allowing them to acquire information faster than if they had to discover the same information independently (Buchsbaum et al., 2011; Harris, 2012; Harris & Koenig, 2006; Jaswal & Kondrad, 2016; Sobel & Kushnir, 2013). However, overreliance on others can lead children to discount or ignore their own observations or hypotheses, a phenomenon termed *testimony bias* (Jaswal, 2010). It appears to be the case both that early learners have a robust capacity to accurately track and structure statistical information from their environments, but also that children will follow what other people say and do, even when input from others conflicts with their own experience and leads to worse outcomes. The testimony bias has been largely examined in children preschool-age and younger; here, we examined this phenomenon in older children who have had more experience with the world and more developed knowledge structures. Our goal was to better understand the conditions under which input from others can enhance or interfere with children's learning.

Research on the testimony bias has primarily focused on infants through preschool-age children. In a classic example (Jaswal, 2010),

Rista C. Plate https://orcid.org/0000-0003-4754-2459
Kristin Shutts https://orcid.org/0000-0002-2323-805X
Aaron Cochrane https://orcid.org/0000-0001-6691-9149
C. Shawn Green https://orcid.org/0000-0002-9290-0262
Seth D. Pollak https://orcid.org/0000-0001-5184-9846

This research was supported by National Institute of Mental Health Grant R01MH61285 to Seth D. Pollak and in part by a core grant to the Waisman Center from the National Institute of Child Health and Human Development (U54 HD090256). Rista C. Plate was supported by a National Science Foundation Graduate Research Fellowship (DGE-1256259) and the Richard L. and Jeanette A. Hoffman Wisconsin Distinguished Graduate Fellowship. We thank the families who participated in this study and the research assistants who helped conduct the research.

The experimental paradigm, deidentified data, and analysis scripts are available on Open Science Framework (https://osf.io/ey4ut/). The studies were not preregistered.

Correspondence concerning this article should be addressed to Rista C. Plate, Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104, United States. Email: rplate@sas.upenn.edu

2- and 3-year-old children were tasked with locating a ball after observing it being dropped into an unfamiliar apparatus that would allow the ball to end in multiple locations. Children overwhelmingly chose the location suggested by an adult despite seeing the ball end up in another location; when asked the location of the ball in the absence of input from an adult, children named the correct location. These types of studies reveal that attending to information from another person may lead children to discount their own sensory experience, preempt children's exploration of alternative hypotheses, or undermine other aspects of learning (Bonawitz et al., 2011; Gweon et al., 2014; Schulz & Bonawitz, 2007). Less is known about how development changes the balance of children's individual learning versus learning from others.

With increasing experience over development, children become more discerning in their use of social information. By 4 to 5 years of age, children favor information provided by those who appear accurate, reliable, knowledgeable, and intelligent, as well as those who have access to relevant information (e.g., Birch et al., 2020; Buchsbaum et al., 2011; Corriveau et al., 2009; Einav & Robinson, 2011; Koenig & Harris, 2005, 2007; Kushnir et al., 2008; Pasquini et al., 2007). Additionally, when children of this age have the opportunity to evaluate another person's reliability against their own observations, children reduce their trust in that person as an informant (Bridgers et al., 2016; Ronfard & Lane, 2018; Ronfard et al., 2017). Therefore, there is strong and consistent data supporting the view that part of what makes children good learners is that they become active in selecting and weighting social and experiential evidence in acquiring knowledge about the world (Hermes et al., 2018, 2019; Shafto et al., 2012; Sobel et al., 2010; Sobel & Kushnir, 2013).

Less attention has been paid to situations in which older children, who typically rely on more rational reasoning, might revert to reliance on social cues, as is typical of younger children. One such situation, which affects people of all ages, is when testimony comes from a group (Asch, 1951; Haun & Tomasello, 2011). When individuals are faced with group testimony that runs counter to their own observations, they tend to defer, even if it means selecting a nonoptimal or incorrect solution. Another more general condition in which children appear to have difficulty choosing between sources of evidence is when there is a high degree of uncertainty. For example, preschool-age children will abandon their chosen label for an ambiguous object (e.g., a pen morphed with a toothbrush) in favor of a label provided by an informant (Li & Yow, 2018). Moreover, when preschool-age children were asked to guess the color of the collar that would appear on different dogs, they were able to use both the base rates of their own experience and the suggestions of an adult witness who was reliable; however, when base rates and testimony conflicted, children discounted their own probabilistic reasoning and followed the informant's recommendation (Gualtieri et al., 2020). In Gualtieri and colleagues' (2020) study, when the informant tended to be inaccurate, children reverted to reliance on their own observations. Thus, because children show selectivity and balancing of information, questions arise regarding developmental changes in children's cognition.

Prior studies have suggested that uncertain outcomes lead preschool-age children to revert to a testimony bias. Here, we expanded previous work by investigating whether uncertainty in social input (in addition to the outcome) influences reliance on testimony. We also examined the stability of this effect in preschool through school-age children. To do so, we adapted a probabilistic learning task (Plate et al., 2018), in which children successfully learned the locations of rewards and maximized reward receipt from probabilistic information. In the present research, participants similarly searched differentially rewarded locations, but prior to selecting, they received nonverbal testimony (i.e., a pointing hand) regarding the location of the reward. Although much research on selective trust has focused on verbal testimony, most commonly in the context of word learning (see Sobel & Finiasz, 2020 for a review and meta-analysis), children often experience nonverbal forms of testimony. For example, technology provides increasing opportunities for people to express advice, ideas, and opinions nonverbally via social media and online learning platforms (e.g., Antheunis et al., 2013; Freberg et al., 2011; Kaplan & Haenlein, 2010). This testimony is often signaled through a visual marker (e.g., a thumbs-up) and can be provided by individuals or groups familiar or unfamiliar to the learner. Additionally, children encounter learning materials with visual markers (e.g., an image of a child who proposes a strategy for solving a problem in a textbook; Riggs et al., 2015, 2017) intended to convey importance or advice (Magner et al., 2014; Rey, 2012).

In Experiment 1, we varied the degree of testimony certainty by manipulating the likelihood that the testimony suggested the correct location of a reward, which allowed us to examine (a) the degree to which participants relied on the testimony, (b) how the reliability of testimony influenced the reward children received, and (c) the strategies children used to obtain rewards. In Experiment 2, we asked whether a nonsocial cue would influence choices similarly to those made based on testimony. In Experiment 3, we introduced uncertainty as a one-time, unexpected change in the reliability of the testimony to explore whether bias to trust in testimony is further influenced by the type of uncertainty (i.e., persistent vs. time-limited). Across experiments, we tested the following hypotheses:

> *Hypothesis 1 (H1):* Under conditions of high uncertainty, children will show testimony bias (i.e., they will over-rely on the testimony at the cost of the reward).

> *Hypothesis 2 (H2):* High uncertainty will elicit testimony bias in older children, even though children as young as 4 years are able to locate reward in this task using probabilistic reasoning.

Experiment 1

Experiment 1 included three between-subjects conditions in which a confederate's testimony indicated a reward location that varied in the level of certainty. Conditions were the *reliable* condition, in which the testimony frequently indicated the location that would be rewarded on each given trial; the *unreliable* condition, in which the testimony infrequently indicated the rewarded location; and the *mixed reliability* condition, in which the testimony indicated the rewarded location on half of the trials. First, we measured overall agreement with the testimony and whether the probability of choosing the location suggested by the testimony changed across trials. Second, we examined whether condition affected participants' ability to obtain rewards. Third, we used a modeling approach to classify the types of strategies children used. Finally, we calculated the weight attributed to these strategies.

We designed both the reliable and unreliable scenarios to present little uncertainty: Either the testimony was a good indicator of the underlying reward, in which case one should follow the suggestions, or it was a poor indicator of the reward, in which case one should ignore the suggestions. However, the mixed reliability condition introduced a higher degree of uncertainty. The relation between the testimony and the reward was not predictable: On any given trial the testimony could suggest the correct location or an incorrect location. This trial-by-trial uncertainty stood in contrast to the reliable and unreliable conditions in which one could be reasonably certain that the testimony would either suggest the correct location (reliable condition) or suggest an incorrect location (unreliable condition). We expected that the mixed reliability condition would present an opportunity to observe the testimony bias. We predicted that participants would continue to select the suggested locations more often than chance (i.e., the frequency with which the confederate suggested the correct location) and, as a result, garner fewer rewards. We tested children ages 4 to 9 years so as to include preschool-age children (for whom there is already evidence of a testimony bias; Gualtieri et al., 2020; Jaswal, 2010; Jaswal et al., 2010) as well as older children.

Method

Participants

Participants were English-speaking children residing in a medium-size city in the midwestern United States. Participants were recruited through a database of families who had previously expressed interest in participating in research studies, flyers/ads placed in the community, or via social media. The sample included 102 children ages 4 to 9 years (the sample included nine Hispanic/Latinx and 93 non-Hispanic/Latinx participants, of whom six were Black/African American, seven were Asian American, eight were multiracial, 77 were White, and four indicated "other" or did not respond). Thirty-six children were in the reliable condition (23 boys, 13 girls; M_{age} = 6.887, SD_{age} = 1.736; see Table S1 in the online supplemental material for demographics for all experiments), 34 children were in the unreliable condition (19 boys, 15 girls; M_{age} = 6.887, SD_{age} = 1.723), and 32 children were in the mixed reliability condition (17 boys, 15 girls; M_{age} = 7.053, SD_{age} = 1.866). The sample size was consistent with previous research using a similar experimental task (Plate et al., 2018). There were no differences in participant age (p = .908) or gender (p = .642) by condition. An additional 16 participants were excluded because they quit the experiment before all trials were presented (n = 13) or because of experimenter error (n = 3). The University of Wisconsin–Madison Institutional Review Board approved the research (Protocol ID: 2018-0520; Title: *Probabilistic Learning Across Development*). Parents of child participants gave informed consent, and children gave verbal assent. Parents received $20 for their time, and children chose a prize after participating.

Procedure

The experimenter brought the participant to the waiting room and explained that the participant would play a computerized game where the goal was to find gold coins hidden behind rocks (adapted from Plate et al., 2018). The participant met a confederate (who, unbeknownst to the participant, was a research assistant in the lab). Most confederates were White women. Participants were told that the confederate had played the game once before and would provide suggestions about which rock to choose throughout the game. The experimenter also explained that the confederate would play the game in another room and would not be able to see the participant's choices. The participant and confederate then proceeded to two separate testing rooms situated across a hallway from each other. We did not include details about the confederate's involvement aside from saying that she would provide "suggestions" to the participant. Our intention was for the context to be ambiguous to make all three conditions of the experiment plausible. For example, if the confederate had an explicit goal of "helping" the participant or had a lot of experience with the task, it may have been particularly surprising, and perhaps less believable, when the confederate frequently suggested unrewarded locations. Similarly, because children are sensitive to myriad cues related to an informant's competence (e.g., Buchsbaum et al., 2011; Corriveau et al., 2009; Einav & Robinson, 2011; Koenig & Harris, 2005, 2007; Kushnir et al., 2008; Pasquini et al., 2007), we limited the participant's interaction with, and information about, the confederate in order for the participant to glean reliability information from the task statistics alone.

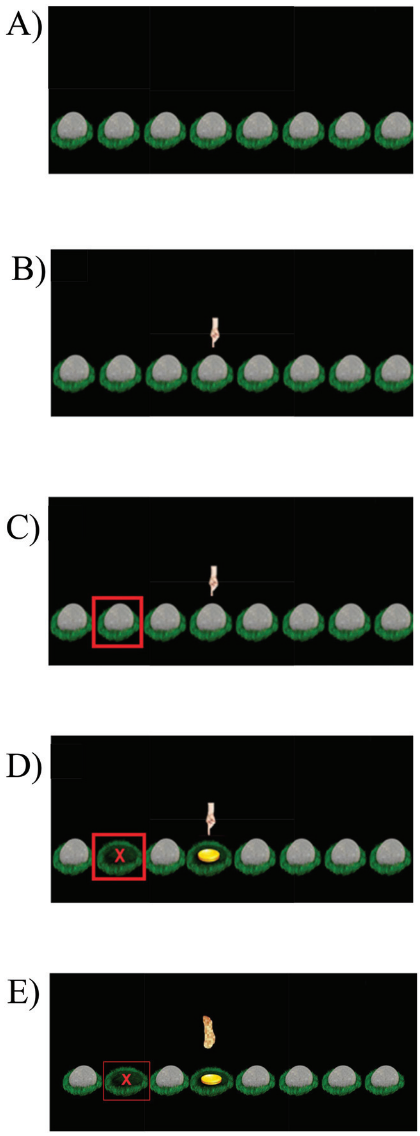
Following a practice phase (see the online supplemental material for details), an icon meant to suggest that the participant's computer was connecting to the confederate's computer appeared on the screen ("Waiting for Player 2. . ."). Additionally, during this time, the experimenter went to the confederate's testing room to "set up." Specifically, the experimenter entered the confederate's room and asked the confederate if they remembered how to play the game (a conversation that could be heard by the child across the hallway with both doors open) before closing the door to the confederate's room. After approximately one minute, the experimenter left the confederate's room saying "good luck" to the confederate. After 30 s, text that read "Player 2 ready. Please wait for experimenter." appeared on the participant's screen. The experimenter returned to the participant's room, wished the participant good luck, and instructed the participant to begin the game by clicking the mouse. The simple set-up proved to make the confederate's involvement in the game believable to participants. None of the children tested indicated doubt about the confederate.

There were two blocks of 100 test trials, separated by a break. At the start of each test trial, eight rocks appeared on the screen with equal spacing along a horizontal line (see Figure 1, Panels A through D for the experimental task). Before participants were allowed to respond, a pointing hand indicated the confederate's suggestion. (The task was programmed such that participants were unable to respond prior to seeing the confederate's suggestion.) When participants selected the correct location on a trial, a coin appeared in place of the rock they selected. When participants selected an incorrect location on a trial, a red "X" appeared in the chosen location and the coin was revealed in the correct location. After completing the task, participants answered questions about the game, the confederate who provided testimony, and the strategy they used during the game (results for participant responses are presented in the online supplemental material).

Design

From left to right, the following probabilities defined the likelihood of a coin appearing at the eight possible rock locations on any given trial (i.e., Location 1 = 0%, Location 2 = 0%, Location 3 = 5%, Location 4 = 10%, Location 5 = 70%, Location 6 = 10%, Location 7 = 5%, Location 8 = 0%). To make all participants'

Figure 1
Progression of Computerized Task



Note. One instance of choice behavior (i.e., choose rock not suggested by confederate) and one possible outcome (i.e., participant fails to find the coin; confederate finds the coin). Panel A: Display prior to participant choice. Panel B: Pointing hand indicating confederate's suggestion. Panel C: Red box highlighting participant's choice. Panel D: Coin appearing in rewarded location. Panel E: The nonsocial cue. See the online article for the color version of this figure.

experiences statistically equivalent, the outcomes were predetermined to ensure a match to the location probabilities across trial blocks (i.e., in each 100-trial block, Rock 5 was rewarded on exactly 70 trials, and Rocks 4 and 6 were rewarded on exactly 10 trials). Participants were not shown the probabilities; the probabilities had to be learned via experience with the task.

The choices of the confederate (i.e., the underlying distribution of the confederate's choices, see Figure S1 in the online supplemental material) varied across three conditions, and participants were randomly assigned to one of these conditions. In the reliable condition, the confederate suggested the correct rock 90% of the time; in the unreliable condition, the confederate suggested the correct rock 10% of the time; in the mixed reliability condition, the confederate suggested the correct rock 50% of the time. Participants were not told about these probabilities. All other task specifications remained constant across conditions. We conducted analyses in R (version 3.6.3; R Core Team, 2019) and used the tidyverse package (Wickham et al., 2019) for data organization, lme4 package (Bates et al., 2012) for linear and mixed effects models, and ggplot2 (Wickham, 2016) for visualization. The experimental task, de-identified data, and analysis script are available on Open Science Framework: https://osf.io/ey4ut/.

Results

Agreement With Testimony

First, we compared participants' overall agreement with the testimony and the testimony's true reliability (see Table 1). We found evidence of a testimony bias in both the unreliable and the mixed reliability conditions, as participants relied on the testimony to a greater extent than the reliability would recommend. Next, we examined participants' likelihood of agreeing with the testimony across condition, age, and time using a logistic mixed effects model. We regressed participant response (1 = followed suggestion, 0 = did not follow suggestion) on the interaction between condition (mixed reliability as referent), age (continuous, mean-centered), time (i.e., trial number, mean-centered and divided by 10 to facilitate convergence) and all lower order effects. We included a by-participant random intercept and a by-participant slope for time. Our reported effects include odds ratios (*OR*), which indicate by how much the odds of agreeing with the confederate's suggestion increase ($OR > 1$) or decrease ($OR < 1$) with each one-unit increase in the independent variable of interest, and we include 95% confidence intervals around the *OR*s. There was an effect of condition ($\chi^2(2) = 409.913$, $p < .001$; see Figure 2). Participants were more likely to follow the suggestions of a reliable, as compared with a mixed reliability, confederate ($b = 2.446$, $p < .001$, $OR = 11.546$, 95% CI [7.291, 18.285]). Participants were also more likely to follow the suggestions of a mixed reliability confederate as compared with an unreliable confederate ($b = 2.276$, $p < .001$, $OR = 9.735$, 95% CI [6.236, 15.197]). This effect is qualified by a condition by time interaction ($\chi^2(2) = 84.310$, $p < .001$; Figure 3 top left). Participants increased their agreement over time in the reliable condition ($b = .200$, $p < .001$, $OR = 1.221$, 95% CI [1.154, 1.291]) and decreased their agreement over time in the unreliable condition ($b = -.050$, $p < .001$, $OR = .951$, 95% CI [.928, .975]); however, agreement did not change over time in the mixed reliability condition ($b = .014$, $p = .395$, $OR = 1.014$, 95% CI [.982, 1.046]). All other effects in the model were not significant.

Reward

We next examined whether uncertain testimony impacted participants' ability to obtain rewards. To do so, we ran the model discussed in the preceding text, with likelihood of finding a gold coin as the dependent variable (1 = found coin, 0 = did not find coin; see the online supplemental material for effects of age and time). Participants found the most coins in the reliable condition, followed by the mixed reliability condition; participants found the

Table 1
Comparison of Overall Participant Agreement With Confederate Reliability for Experiments 1 and 2

Condition	Testimony reliability	Participant agreement	t[a]	*df*	*p*	Cohen's *d*	95% CI
Experiment 1							
Unreliable	10%	21%	6.696	33	<.001	1.15	[18, 24]
Mixed reliability	50%	65%	4.8	31	<.001	0.85	[59, 71]
Reliable	90%	92%	1.652	35	.108	0.28	[90, 95]
Experiment 2							
Unreliable	10%	19%	10.49	28	<.001	1.95	[17, 20]
Mixed reliability	50%	52%	0.73	28	.472	0.14	[46, 59]
Reliable	90%	86%	−1.354	30	.186	−0.24	[79, 92]

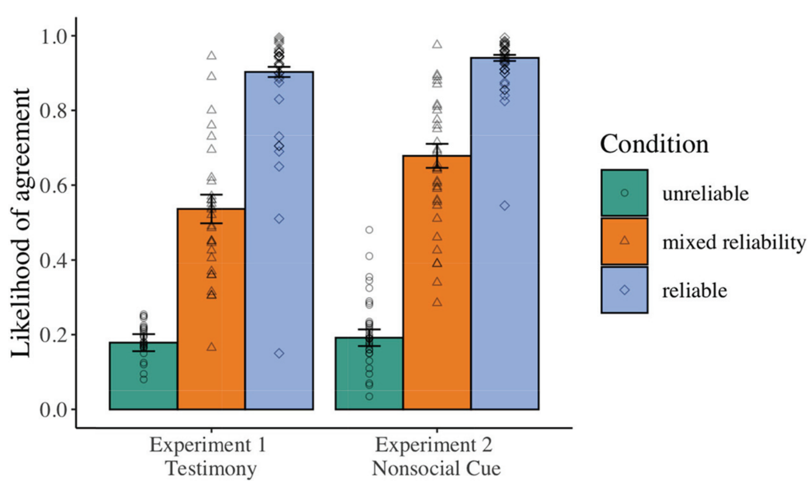
Note. A post hoc sensitivity analysis using G*Power (Faul et al., 2007) for a one-tailed t test, error probably of .05, power of .80, and sample size of 29 (i.e., the smallest group in the study) reveals a critical t value of 1.70 and effect size d of .47.
[a] Participant agreement versus testimony reliability.

fewest coins in the unreliable condition ($\chi^2(2) = 487.603$, $p < .001$; reliable versus mixed reliability condition $b = 1.810$, $p < .001$, $OR = 6.113$, 95% CI [5.005, 7.468]; mixed reliability versus unreliable condition $b = .239$, $p = .018$, $OR = 1.270$, 95% CI [1.042, 1.547]). However, the main effects were qualified by a condition by time interaction ($\chi^2(2) = 21.703$, $p < .001$; see Figure 3, bottom left). Although participants in all three conditions improved their reward receipt over time (reliable: $b = .066$, $p < .001$, $OR = 1.068$, 95% CI [1.046, 1.092]; mixed reliability: $b = .021$, $p < .001$, $OR = 1.021$, 95% CI [1.011, 1.032]; unreliable: $b = .062$, $p < .001$, $OR = 1.064$, 95% CI [1.046, 1.082]), there were relatively greater gains over time in the reliable and unreliable conditions as compared with the mixed reliability condition (*b*s = .043 and .041, *p*s < .001; $OR = 1.044$, 95% CI [1.022, 1.066] and $OR = 1.042$, 95% CI [1.021, 1.063], respectively). These results converge with the results measuring agreement with the confederate's suggestions. Participants are less able to increase their reward receipt over time considering uncertain testimony.

Children's Use of Strategies

To better characterize response patterns, we examined the strategies participants used. In past research, participants who completed this computer task without testimony transitioned over time from a suboptimal strategy (i.e., probability matching = choosing each option at the rate it is rewarded) to an optimal strategy (i.e., maximizing = choosing the most highly rewarded option; Plate et al., 2018). Here, we were interested in participant strategy use in the presence of testimony. Therefore, we assessed the extent to which individual participant choices were best captured by one of five different possible models of choice behavior. In brief, the first model was a probability-matching model. Here participants were expected to choose each option in proportion to the probability that each location had been observed to be correct up to the current trial. The second model was a maximizing model. Under this model, participants were expected to choose the option that had been observed to have the highest probability of reward up to the current trial. The third model was a testimony-matching model, in which the participant's distribution of choices was expected to be the same as the confederate's distribution of choices. The fourth model was a testimony-following model, in which participants were expected to choose the option suggested by the confederate (this model can be thought of as maximizing on the confederate's suggestions). The final model was a random choice model, in which there was an equal and constant probability of the

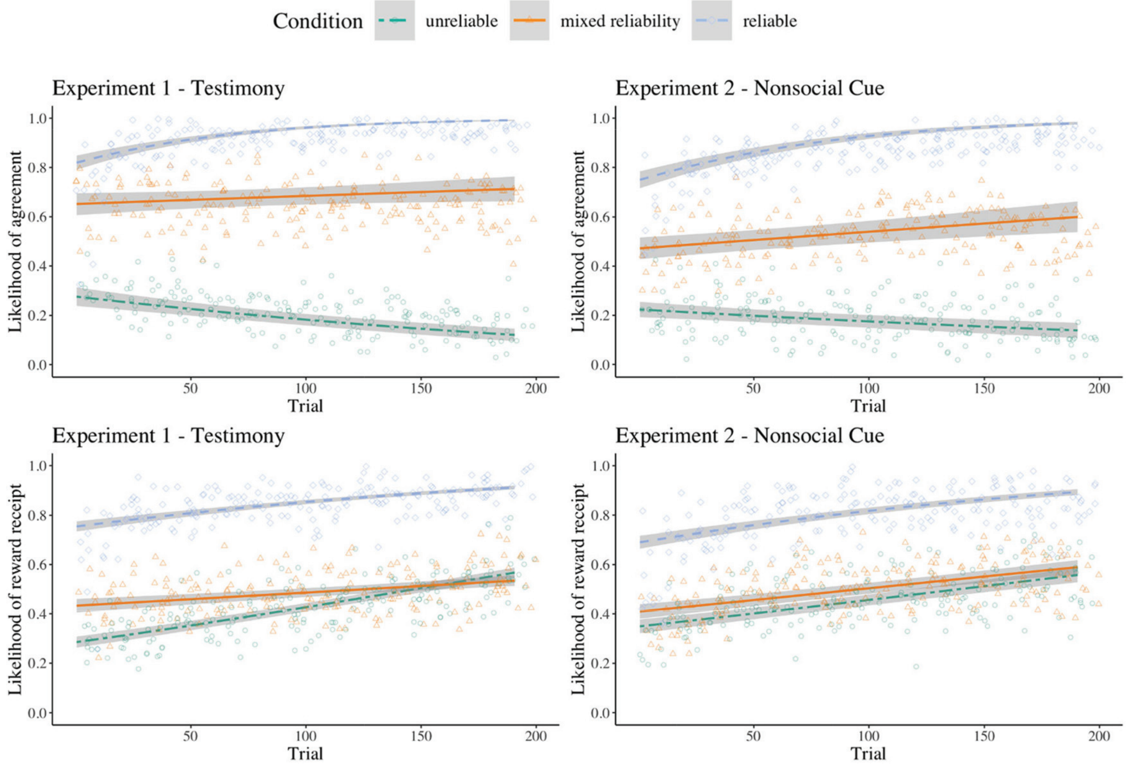
Figure 2
Likelihood of Agreeing With the Testimony/Cue by Condition and Experiment



Note. Model predictions and participant-level data. Error bands represent standard error of the estimates. Points represent individual participants averaged across trials. See the online article for the color version of this figure.

Figure 3

Likelihood of Agreeing With the Testimony/Cue (Top Row) and the Likelihood of Reward (Bottom Row) for Each Experiment 1 and 2



Note. Model predictions and participant-level data. Lines are point estimates from linear model with the interaction between condition and trial number, and lower-order effects. Error bands represent standard error of the estimates. Points are responses averaged across participants. See the online article for the color version of this figure.

participant selecting each of the eight options. (For full model details, see the online supplemental material.)

To address what approaches children across the different conditions used, we measured participants' use of the strategies described in the preceding text. To do this, we determined the log likelihood of each participant's set of choices across the experiment by summing the log probabilities of the choices given the probability structure across rock locations and testimony considering each participant's past trial experience and the model being evaluated. We regressed the log likelihood on condition (mixed reliability condition set as the referent), model (testimony following as referent), age (continuous, mean-centered), and all possible interactions with a by-participant random intercept.

There was an interaction between model (probability matching, maximizing, testimony following, testimony matching, and random choice) and condition ($F(8, 396) = 117.557$, $p < .001$; see Table S2 in the online supplemental material for all pairwise comparisons and Table S3 for the percentage of participants best fit by each strategy). Most participants in the reliable condition were best fit by the testimony-following model (97%), which was the optimal strategy for obtaining reward in this condition. Most participants in the unreliable condition were best fit by the probability-matching model (85%), suggesting that they were making choices based on the underlying reward distribution (although they were not using the optimal strategy of maximizing given the reward distribution). Finally, the participants in the mixed reliability condition were best fit by either the probability-matching (44%) or the testimony-following (50%) model. In the condition with the most uncertainty, participants were not using an optimal strategy (maximizing). Moreover, half of the participants were simply following testimony.

Weighting of Testimony

To better understand how participants were using each of these strategies, we examined a mixture model that included both probability matching and testimony following because these were overwhelmingly the strategies used by participants. One key aspect of the mixture model was that the mixture itself was not constrained to be a perfect average of the given two models at hand (i.e., 50% probability matching, 50% testimony following). Instead, the fitting procedure involved finding the best-fitting weighted average of the two models (e.g., if a participant's choices were largely consistent with testimony following, but occasional choices were more consistent with probability matching, this might produce a final mixture with weights of 90% testimony following, 10% probability matching). Examining the specific weight participants attributed to the testimony-following model versus the probability-matching model thus provided critical insight into how children used the testimony and underlying reward cues.

We regressed the weight attributed to the testimony-following model on condition (mixed reliability as referent), age (continuous, mean-centered), and their interaction. When comparing the weight given to the testimony-following model against that given to the probability-matching model within the mixture model that considered both, there were effects of condition, $F(2, 96) = 181.981$, $p < .001$, $\Delta R^2 = .776$, and age, $F(1, 96) = 9.179$, $p = .003$, $\Delta R^2 = .020$, which were qualified by a marginal condition-by-age interaction ($F(2, 96) = 2.994$, $p = .055$, $\Delta R^2 = .013$; see Figure 4). A sensitivity analysis for a linear multiple regression using G*Power (Faul et al., 2007) with $\alpha = .05$, power = .80, sample size = 102, and number of predictors (5) specifies a critical F value of 2.31 and effect size f^2 of .13; the critical F value was exceeded for the effects described above, f^2 was exceeded for the effect of condition (condition, $f^2 = 3.79$; age, $f^2 = .10$; condition by age, $f^2 = .06$). Examining simple slopes for the relation between participant age and weight attributed to the testimony-following model, participant age mattered in the mixed reliability condition: weight attributed to the testimony decreased with age (unreliable: $b = -.009$, $p = .629$; mixed reliability: $b = -.051$, $p = .003$; reliable: $b = .005$, $p = .759$). Therefore, younger participants had more difficulty disengaging from the testimony under uncertainty. In contrast, participants in the unreliable condition attributed little weight to the testimony, regardless of age. As would be expected based on strategy use, participants in the reliable condition attributed great weight to the testimony, again regardless of age.

Discussion

Experiment 1 provides evidence for testimony bias in middle childhood. Across the full sample, children over agreed with the testimony both when the confederate was mixed in terms of reliability and when the confederate was unreliable. However, when the testimony predictably provided poor recommendations (unreliable condition), children reduced their trust over time, resulting in greater improvement of reward receipt over time. On the contrary, participants facing a higher degree of uncertainty in the testimony (mixed reliability condition) did not change their use of that cue over time and improved relatively less in their reward receipt over time. This pattern of results points to the idea of uncertainty not just influencing overall choices, but also influencing the learner's ability to flexibly adjust choices over time.

We did not find age differences in likelihood of agreeing with the testimony. We did, however, find that children attributed less weight to the testimony-following strategy with age. That children attributed less weight to the testimony-following strategy with age provides a hint of how older children could be working to overcome a testimony bias. Namely, by first moving away from a strategy that is reliant on input from another. Nevertheless, participants across the age range continued to be disproportionately influenced—to some degree—by the testimony. One piece of supporting evidence for this idea is the finding that participants did not use the optimal, maximizing, strategy. Participants in the unreliable condition were best fit by the probability-matching model—not the maximizing model. The only condition in which children demonstrated optimal choice behavior was when receiving testimony that was highly reliable (note that because the reliable testimony was correct 90% of the time, always following this testimony was a better strategy than employing the maximizing model, which would be correct 70% of the time). Taken together, the results from Experiment 1 indicate that uncertainty affected children's ability to disengage from testimony and subsequently enact optimal choice behavior.

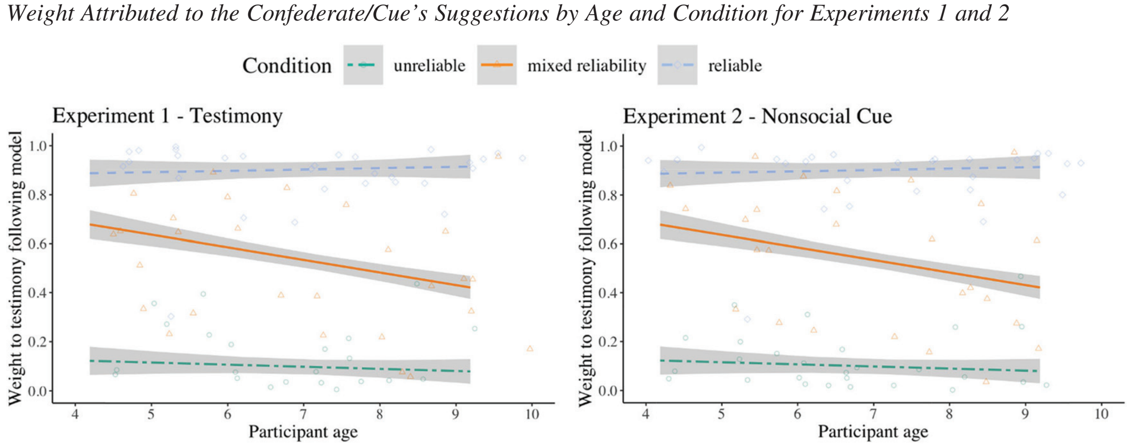
One account of participants' performance is that they were operating under a testimony bias. Another possibility, however, is that children's performance in Experiment 1 is the result of having two cues (the testimony and the rewarded location) to track. Tracking multiple cues may create difficulty for discerning the best strategy more generally. Therefore, we conducted Experiment 2 to better understand how children consider cues that are not social.

Experiment 2

Experiment 1 provided evidence consistent with a testimony bias (i.e., participants overrelying on the suggestions from the confederate) in a condition with high uncertainty. However, in order to have confidence that the pattern of behavior was a result of receiving testimony from another person, we replicated the method

Figure 4

Weight Attributed to the Confederate/Cue's Suggestions by Age and Condition for Experiments 1 and 2



Note. Model predictions and participant-level data. Lines are point estimates from linear model with the interaction between condition and participant age, and lower order effects. Error bands represent standard error of the estimates. Points are individual participants' weights. See the online article for the color version of this figure.

of Experiment 1, with the exception that we provided "testimony" (i.e., a cue) that was not associated with a social agent. We predicted that fewer participants would continue to follow the cue in the mixed reliability condition of Experiment 2 (compared with Experiment 1). Such a finding would be in line with research showing that infants use social, but not nonsocial, cues in uncertain learning situations (Tummeltshammer et al., 2014). We were also interested in whether children would continue to probability match, instead of maximize, in the unreliable condition. It may be that children would more readily disregard a nonsocial cue, which might facilitate more time to transition to maximizing.

Method

Participants

The sample included 89 children ages 4 to 9 years (five Hispanic/Latinx, 80 non-Hispanic/Latinx, of whom one was Black/African American, three were Asian American, 15 were multiracial, 66 were White, and four did not report race or ethnicity). Children were recruited from the same community as Experiment 1; none of the children had participated in Experiment 1. Thirty-one children were in the reliable condition (19 boys, 12 girls, $M_{\text{age}} = 7.118$, $SD_{\text{age}} = 1.779$), 29 children were in the unreliable condition (17 boys, 12 girls, $M_{\text{age}} = 6.898$, $SD_{\text{age}} = 1.671$), and 29 children were in the mixed reliability condition (16 boys, 13 girls, $M_{\text{age}} = 6.986$, $SD_{\text{age}} = 1.577$). There were no differences in child age ($p = .877$) or gender ($p = .891$) by condition. An additional 17 participants were excluded because they ended the experiment before all trials were presented. The Institutional Review Board approved the research. Parents of child participants gave informed consent; children gave verbal assent. Children chose a prize after participating.

Procedure

The procedure was identical to Experiment 1 with the following exceptions. The cue signaling the suggested rock was an image of neutral-colored pixels (see Figure 1, Panel E) arranged to approximate the shape and size of the pointing hand from Experiment 1. Second, we revised the instructions of the task to remove any suggestion that the cue was social. We told participants, "You might notice that sometimes you can see these funny shapes. They're from the computer, and I'm not sure why they're showing up."

The reasoning behind the change in the instructions was to emphasize the nonsocial nature of the cue. There are difficulties in designing paradigms to contrast social and nonsocial information, particularly in probabilistic learning tasks. Perhaps most difficult is that there is a tendency for individuals to view patterned behavior as social in nature (Heider & Simmel, 1944). Further, children have increasing knowledge about, and interactions with, technology that might bias them to see a computer cue as having some social characteristics. For example, children in kindergarten and second grade do not readily differentiate between teachers and information from the Internet when seeking knowledge (Wang et al., 2019). To make the present paradigm convincingly nonsocial, we devised an explanation for the cue that discouraged social attributions.

Results

Agreement With the Cue

In comparing participants' overall agreement with the cue with the cue's reliability, participants only over-agreed with the cue in the unreliable condition (see Table 1). From the logistic regression, there was an effect of condition ($\chi^2(2) = 257.195$, $p < .001$, reliable versus mixed reliability: $b = 2.392$, $p < .001$, $OR = 10.928$; 95% CI [6.623, 18.033], unreliable versus mixed reliability: $b = -1.709$, $p < .001$, $OR = .181$, 95% CI [.110, .297]; see Figure 2), and an effect of time ($b = .027$, $\chi^2(1) = 3.758$, $p = .053$, $OR = 1.028$, 95% CI [1.000, 1.056]), which were qualified by a condition-by-time interaction, $\chi^2(2) = 66.538$, $p < .001$ (see Figure 3, top right). There was a relatively greater change in agreement over time for the reliable and unreliable conditions as compared with the mixed reliability condition (*b*s = .119, –.058, *OR*s = 1.126, 95% CI [1.080, 1.175], and .944, 95% CI [.907, .982], $ps < .005$, respectively). Taken together, testimony bias was not present in the condition with the highest uncertainty (the mixed reliability condition), however, participants still failed to update their behavior over time in this condition.

Reward

The effects of condition ($\chi^2(2) = 206.013$, $p < .001$, reliable versus mixed reliability: $b = 1.48$, $p < .001$, $OR = 4.392$, 95% CI [3.423, 5.635], unreliable versus mixed reliability: $b = -.189$, $p = .137$, $OR = .827$, 95% CI [.645, 1.062]), time ($b = .038$, $\chi^2(1) = 20.554$, $p < .001$, $OR = 1.039$, 95% CI [1.022, 1.056]), a condition by time interaction ($\chi^2(2) = 6.924$, p .031, reliable versus mixed reliability: $b = .031$, $p = .011$, $OR = 1.032$, 95% CI [1.007, 1.057], unreliable versus mixed reliability: $b = .007$, $p = .585$, $OR = 1.007$, 95% CI [.983, 1.030]; see Figure 3, bottom right) were consistent with the findings discussed in the preceding text. Additionally, there was a condition by time by age interaction, $\chi^2(2) = 11.293$, $p = .004$. Whereas children across ages improved to receive high reward in the reliable condition, children showed more marked improvement with age in those conditions (reliable vs. mixed reliability: $b = -.018$, $p = .017$, $OR = .982$, 95% CI [.968, .997]; reliable vs. unreliable: $b = -.024$, $p = .001$, $OR = .977$, 95% CI [.963, .991]). Therefore, older children show less evidence of a testimony bias influencing reward receipt when a cue is not social in nature.

Children's Use of Strategies

We assessed the same five strategies as in Experiment 1. Most participants in the reliable condition were best fit by the testimony-following model (81%). Most participants in the unreliable condition were best fit by the probability-matching model (83%). Contrary to Experiment 1, the majority of participants in the mixed reliability condition were also best fit by the probability-matching model (69%; 21% fit by the testimony-following model; omnibus interaction between model and condition $F(8, 344) = 72.252$, $p < .001$, pairwise comparisons in the online supplemental material), again suggesting reduced reliance on the cue.

Weighting of the Cue

There was an effect of condition such that participants attributed more weight to the cue in the reliable condition and less weight to the cue in the unreliable condition compared with the mixed reliability condition (omnibus: $F(2, 83) = 78.293$, $p < .001$, $\Delta R^2 = .639$; unreliable vs. mixed reliability: $b = -.273$, $p < .001$; unreliable vs. reliable: $b = -.734$, $p < .001$; reliable vs. mixed reliability: $b = -.460$, $p < .001$; see Figure 4). There was no effect of age and no age by condition interaction (ps $> .1$). A sensitivity analysis for a linear multiple regression using G*Power (Faul et al., 2007) with $\alpha = .05$, power = .80, sample size = 89, and number of predictors = 5 specifies a critical F value of 2.32 and effect size f^2 of .15; the critical F value and f^2 were exceeded for the effect of condition ($f^2 = 1.88$).

Comparing Experiments 1 and 2

To better unpack the differences in how the social testimony and nonsocial cue influenced participant behavior on the task, we directly compared Experiments 1 and 2. There were no differences in participant age ($p = .795$) or gender ($p = 1$) by experiment.

Agreement With the Testimony/Cue. We regressed likelihood of choosing the suggested location on condition, age, and experiment (nonsocial = –.5, social = .5) with by-participant and by-time random intercepts. There was a main effect of experiment ($b = .603$, $\chi^2(1) = 8.065$, $p = .005$, $OR = 1.827$, 95% CI [1.205, 2.769]), with overall greater likelihood of taking the suggestion when the cue was supposedly coming from another person (i.e., testimony in Experiment 1). There was also an effect of condition ($\chi^2(2) = 700.206$, $p < .001$; reliable vs. mixed reliability: $b = 2.053$, $p < .001$, $OR = 7.788$, 95% CI [5.797, 10.464]; unreliable vs. mixed reliability: $b = -1.927$, $p < .001$, $OR = 0.146$, 95% CI [0.108, 1.195]) and a condition by age interaction, $\chi^2(2) = 8.646$, $p = .013$. The interaction pattern shows that with increasing age, participants better distinguished between the reliable and mixed reliability conditions ($b = .255$, $p = .004$, $OR = 1.290$, 95% CI [1.084, 1.534]). Overall, these results point to a testimony bias that is stronger to social testimony (as opposed to a nonsocial cue).

Reward. In addition to main effects of condition ($\chi^2(2) = 619.984$, $p < .001$; reliable vs. mixed reliability: $b = 1.620$, $p < .001$, $OR = 5.056$, 95% CI [4.315, 5.923]; unreliable vs. mixed reliability: $b = -0.211$, $p = .009$, $OR = 0.809$, 95% CI [0.691, 0.948]) and age ($b = 0.073$, $\chi^2(1) = 4.533$, $p = .033$, $OR = 1.076$, 95% CI [1.006, 1.151]) there was a condition by experiment interaction, $\chi^2(2) = 6.961$, $p = .031$. Participants earned relatively more reward in the reliable condition (vs. mixed reliability, $b = .346$, $p = .032$, $OR = 1.414$, 95% CI [1.030, 1.940], vs. unreliable, $b = .386$, $p = .016$, $OR = 1.471$, 95% CI [1.075, 2.014]) in Experiment 1 (social testimony) as compared with Experiment 2 (nonsocial cue). In other words, it appears that overreliance on the testimony (as opposed to the nonsocial cue's placement) in the unreliable and mixed reliability conditions negatively impacted reward receipt.

Strategy Use. We compared the proportion of participants who used the testimony-following model in each condition across experiments. Exactly zero participants used the testimony-following strategy in the unreliable condition in both experiments. Marginally, a greater proportion of participant were best fit by the testimony-following model in the reliable condition of Experiment 1 versus Experiment 2, $\chi^2(1) = 3.281$, $p = .070$. A significantly greater proportion of participants used the testimony-following strategy in the mixed reliability condition with testimony, compared with the nonsocial cue, $\chi^2(1) = 4.468$, $p = .034$, providing evidence of a testimony bias under uncertainty that is particularly robust in response to a social cue.

Weighting of Testimony/Cue. To compare the weighting of strategies, we regressed the weight of the testimony-following model (within the mixture model including testimony following and probability matching) on age, condition, experiment, and all interactions. Several effects emerged. In addition to main effects of age ($b = -.046$, $F(1, 179) = 8.774$, $p = .003$, $\Delta R^2 = .013$) and condition (omnibus: $F(2, 179) = 232.446$, $p < .001$, $\Delta R^2 = .682$; unreliable vs. mixed reliability: $b = -.355$, $p < .001$; reliable vs. mixed reliability: $b = .414$, $p < .001$) that were consistent with Experiment 1, there was a main effect of experiment, again suggesting that more weight was attributed to testimony as compared with the nonsocial cue ($b = .196$, $F(1, 179) = 14.191$, $p < .001$, $\Delta R^2 = .021$). The main effects were qualified by a condition by age interaction suggesting that participants decreased the weight they attributed to the suggestions in the mixed reliability condition with age, but age did not influence behavior in the unreliable or reliable conditions (omnibus: $F(2, 179) = 4.489$, $p = .013$, $\Delta R^2 = .013$; simple slopes: unreliable [$b = -.008$, $p = .615$], mixed reliability [$b = -.046$ $p = .003$], reliable [$b = -.017$, $p = .239$]). Finally, though not statistically significant, there was a marginal condition by experiment interaction, $F(2, 179) = 2.474$, $p = .087$, $\Delta R^2 = .007$, reflecting the pattern such that there were greater differences in weight attributed to testimony versus the nonsocial cue in the mixed reliability condition. A sensitivity analysis for a linear multiple regression using G*Power (Faul et al., 2007) with $\alpha = .05$, power = .80, sample size = 191, and number of predictors = 11 specifies a critical F value of 1.84 and effect size f^2 of .09; the critical F value was exceeded for the effects of condition, age, and experiment and the interactions between condition and age and condition and experiment. The f^2 was exceeded for the effect of condition ($f^2 = 2.60$; age $f^2 = .05$, experiment $f^2 = .08$, condition by age, $f^2 = .05$, condition by experiment, $f^2 = .03$).

Discussion

In general, participants were more likely to select a suggested location in the social experiment as compared with when the cue was nonsocial. Particularly when the cue was mixed in terms of its reliability, children attributed more weight to the testimony and used a strategy that relied on this information (i.e., they followed the confederate's suggestions). When the cue was not social, children primarily focused on the underlying reward. Therefore, the distinction between the testimony and nonsocial cues was contingent on the uncertainty of the environment.

The idea that testimony guides behavior under uncertainty whereas nonsocial cues do not is consistent with research with infants (Tummeltshammer et al., 2014). Specifically, infants directed their attention to a never-rewarded location on a computer screen when that location was cued by a person, but not when it was cued by an arrow. Additionally, the results from the current research suggest that the ability to disengage from a social cue in order to obtain higher reward increases with age. Notably, this age-related finding was absent in the nonsocial experiment, providing a hint that testimony's effect may be present later across development. In all, the comparison with a nonsocial experiment provides support for the idea that there is a lingering effect of a testimony bias that appears to be—at least in part—unique to uncertain social information.

Experiment 3

When the testimony was characterized by trial-by-trial uncertainty, as in the mixed reliability condition of Experiment 1, weight to the testimony-following strategy decreased with age. A comparison with a nonsocial cue provided evidence that the social nature of the uncertain cue impacted strategy use. Yet, there may be more than one way in which testimony can present uncertainty, for example testimony can change unexpectedly. In Experiment 3, we presented conditions in which testimony unexpectedly changed from being reliable to being unreliable or vice versa. In general, we expected that children would be able to flexibly update their behavior in this changing social environment, particularly given how well distinguished the conditions of Experiment 1 were in terms of agreement with the testimony and because children as young as 3 years of age will decrease their trust in verbal testimony after a seeing evidence that an informant's claim was incorrect (Hermansen et al., 2021). We were primarily interested in whether age might influence flexibility. In Experiment 1, weight to the testimony-following model decreased with age when the testimony only sometimes indicated the correct reward location. Additionally, adults were better at detecting a one-time pattern change in a probability-learning task than 4–six-year-old children (Starling et al., 2018). Further, individuals use minimal information to make trait-level attributions about social agents (Uleman et al., 2008), and children who are younger than 5 years of age are less accurate when making predictions about a social agent's future behavior because they are not able to readily incorporate consistency cues (Boseovski & Lee, 2006). Based on preschool-age children's difficulty adapting to changing information and their bias to trust in testimony, we reasoned that younger children might attribute more weight to testimony that is reliable early in the experiment even if that reliability changes over time. We predicted that younger children would be more likely to follow the testimony if the confederate had previously reliably indicated the rewarded location, even if that confederate currently provided unreliable suggestions. Such a difference may not be present when interacting with initially unreliable testimony. Experiment 2 suggests that children will monitor information sources, even when they have no reason to expect that they will be useful. Therefore, we predicted that, even if participants did not use the testimony in the first half of the experiment, they would be sensitive to the change in reliability and update their choices accordingly.

Method

Participants included 129 children ages 4 to 9 (12 Hispanic/Latinx, 117 non-Hispanic/Latinx, of whom 10 were Black/African American, four were Asian American, nine were multiracial, 103 were White, and three chose "other" or did not report race) who did not participate in Experiments 1 or 2. Nine additional participants were excluded for not completing the task ($n = 6$) or experimenter error ($n = 3$).

The design, procedure, and data analysis approach were the same as Experiment 1, with the following exceptions: First, in Experiment 3, the conditions included reliable→unreliable (i.e., testimony indicated the correct rock 90% of the time for the first 100 trials and 10% of the time for the second 100 trials, $N = 63$, 31 boys, 32 girls, $M_{\text{age}} = 6.938$, $SD_{\text{age}} = 1.691$) and unreliable→reliable, that is, testimony indicated the correct rock 10% of the time for the first 100 trials and 90% of the time for the second 100 trials ($N = 66$, 37 boys, 29 girls; $M_{\text{age}} = 7.068$, $SD_{\text{age}} = 1.630$). There was no difference in participant age ($p = .531$) or gender ($p = .328$) by condition. There was no break or any other change in the experiment that would draw participant's attention to the change in reliability. As in Experiments 1 and 2, we examined likelihood of choosing the suggested location, likelihood of finding a coin, participant strategy use, and weight attributed to strategies.

Results

Agreement With Testimony

We regressed whether the participant chose the rock that was indicated by the testimony on condition (unreliable→reliable = –.5, reliable→unreliable = .5), age (continuous, mean-centered), task half (first half = –.5, second half = .5), and all possible interactions. We included a by-participant random slope for task half and included a by-trial random intercept. The interaction between condition and age was significant ($b = -.126$, $\chi^2(1) = 41.430$, $p < .001$, $OR = .881$, 95% CI [.848, .916]; observed power via *simr* (Green & MacLeod, 2016) was 100.00%, 95% CI [98.17, 100.00]; see Figure 5, Panel A). For participants in the reliable→unreliable condition, likelihood of agreeing with the testimony decreased with age ($b = -.061$, $p < .001$, $OR = .941$, 95% CI [.915, .967]), whereas in the unreliable→reliable condition, the likelihood of agreeing with the testimony increased with age ($b = .073$, $p < .001$, $OR = 1.075$, 95% CI [1.046, 1.106]). In sum, younger participants were more strongly influenced by the initial testimony both in the reliable→unreliable and unreliable→reliable conditions.

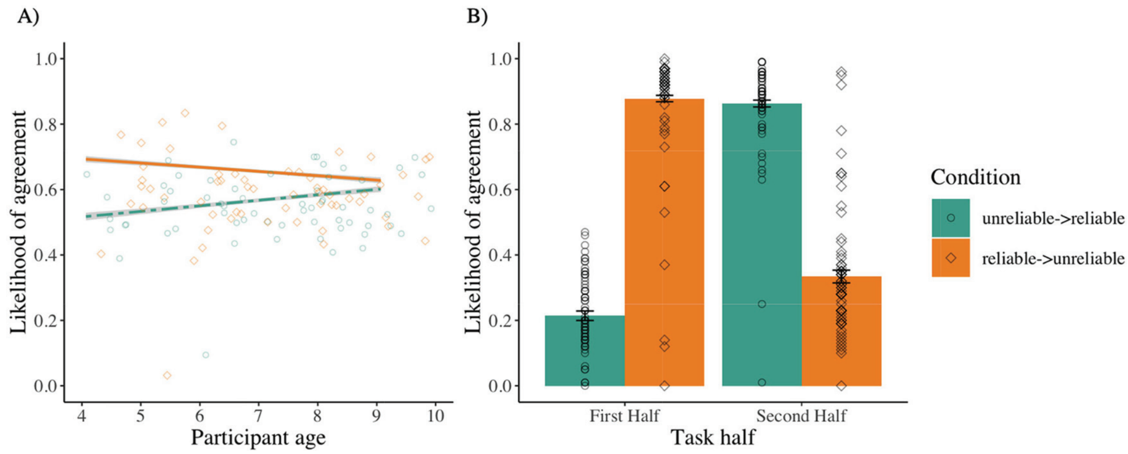
In addition to the condition-by-age interaction, there was a main effect of condition ($b = .373$, $\chi^2(1) = 122.018$, $p < .001$, $OR = 1.452$, 95% CI [1.359, 1.551]) and condition by half interaction ($b = -5.805$, $\chi^2(1) = 586.893$, $p < .001$, $OR = .003$, 95% CI [.002, .005]; see Figure 5, Panel B), indicating greater differences between conditions in the first, as compared with the second, half of the experiment. There was also a main effect of half ($b = .237$, $\chi^2(1) = 3.877$, $p = .049$, $OR = 1.268$, 95% CI [1.001, 1.606], the greater likelihood of agreement with testimony in the second half of the experiment).

Reward

There were numerous effects on reward. First, a main effect of condition ($b = .080$, $\chi^2(1) = 7.392$, $p = .007$, $OR = 1.084$, 95% CI [1.023, 1.148]), indicated that participants obtained more reward when the testimony was initially reliable. A main effect of age ($b = .102$, $\chi^2(1) = 130.898$, $p < .001$, $OR = 1.108$, 95% CI [1.088, 1.127]), showed that reward receipt increased with age. The main effects were qualified by condition by age, ($b = -.038$, $\chi^2(1) = 4.523$, $p = .033$, $OR = .963$, 95% CI [.930, .997]), and condition by half interactions ($b = -3.726$, $\chi^2(1) = 390.577$, $p < .001$, $OR = .024$, 95% CI [.017, .035]). The condition by age interaction suggests that there was a relatively smaller difference between conditions in likelihood of finding coins across age. The condition by half interaction shows greater differences between conditions in the first, as compared with the second, half of the experiment.

Figure 5

Likelihood of Agreement With the Confederate's Suggestion by Age (Panel A) and Task Half (Panel B) for Experiment 3



Note. Model predictions and participant-level data. Lines are point estimates from logistic mixed-effects models with the interaction between condition and participant age, and lower order effects. Error bands represent standard error of the point estimates. Points are individual participants' proportion of choices that followed the confederate's suggestion by age (Panel A) or task half (Panel B). Panel A: Likelihood of agreeing with the confederate's suggestion by age and condition in Experiment 3. Panel B: Likelihood of agreeing with the confederate's suggestion by task half and condition in Experiment 3. See the online article for the color version of this figure.

Children's Use of Strategies

We regressed the log likelihood fit on condition, model, task half, and possible interactions. We included by-participant random slopes for model and task half. The three-way interaction was significant, $F(4, 508) = 731.109$, $p < .001$. Prior to the switch in testimony reliability, most participants who received reliable testimony were best fit by the testimony-following model (92%), whereas participants who received unreliable testimony were best fit by the probability-matching model (79%). Post switch, participants changed their strategy. Most participants who initially received reliable testimony, but now received unreliable testimony, were best fit by the probability-matching model (68%). Participants who initially received unreliable testimony, but now received reliable testimony were best fit by the testimony-following model (80%). When comparing participants who received reliable testimony either in the first half or second half of the experiment, there was no difference in the proportion of participants who were best fit by the testimony-following model ($p = .094$); similarly, no difference in proportion of participants receiving unreliable testimony fit by the probability matching model based on the first or second half of the experiment ($p = .247$).

Weighting of Testimony With an Unexpected Change

Finally, we examined the weight attributed to the testimony-following model and tested whether weight was influenced by initial experience (i.e., receiving reliable testimony initially or receiving unreliable testimony initially) and age. There was a main effect of condition ($b = .097$, $F(1, 125) = 12.568$, $p < .001$; greater weight was attributed to the testimony-following model in the reliable→unreliable condition). This effect was qualified by a condition by half interaction ($b = -1.20$, $F(1, 125) = 485.318$, $p < .001$) consistent with the pattern described in the preceding text. No other effects or interactions were significant ($ps > .1$). Because task half is a repeated measure, we used *simr* (Green & MacLeod, 2016) to estimate the *minimum effect size* (defined as the unstandardized coefficient) to achieve 80% power, which was .078 (power = 83.50%, 95% CI [77.62, 88.36]) for the effect of condition (actual condition coefficient = .097; observed power for effect of condition = 97.00%, 95% CI [93.58, 98.89]) and .16 (power = 82.50%, 95% CI [76.51, 87.50]) for the interaction between condition and half (actual condition by half, coefficient = –1.20; observed power for effect of condition by half = 100.00%, 95% CI [98.17, 100.00]).

Discussion

Here we tested uncertainty, which we defined as an unexpected change in testimony. Whether encountering reliable testimony following a period of unreliability or vice versa, participants readily adjusted strategies to match the uncertain, changing environment. However, younger participants continued to be influenced by the initial testimony as revealed by the difference in young children's likelihood of agreeing with the testimony by condition (an effect that decreased with age). Therefore, younger participants again appeared to be more influenced by the uncertainty—as was also seen in the mixed reliability condition of Experiment 1—and had difficulty adapting to the changing environment as compared with older children.

The influence of initially unreliable testimony also decreased with age. There is some research to support the idea that younger children are more influenced by inaccuracy, though this research suggests that children should be sensitive to the degree of inaccuracy by 4 years of age (Pasquini et al., 2007) and others have found children as young as 4 years to be sensitive to uncertainty (Mcloughlin et al., 2021). However, it is possible that the complexity of the task (e.g., multiple probabilities to track across many trials) prolonged the developmental trajectory. Together, these results hint at the possibility that young children have a more general bias to be influenced by initial testimony—be it reliable or unreliable. However, the age findings should be interpreted with caution for two reasons. First, the effect sizes are small, warranting replication with sample sizes chosen a priori that provide sensitivity to test for the effect. Second, without a nonsocial comparison, we cannot assert that these patterns are caused by the social nature of the cue. Because age-related differences in flexible updating have been found in nonsocial contexts (Starling et al., 2018), it is possible that age-related differences are primarily driven by executive function, memory, or other cognitive skills that track with age. Nevertheless, Experiment 3 provides evidence that children can and do update their choices and strategies in the face of a one-time change in testimony.

General Discussion

The aim of the present research was to investigate whether under conditions of high uncertainty, children would show testimony bias (H1) and whether, even though children as young as 4 years of age can solve this task on their own using probabilistic reasoning, high uncertainty would elicit the testimony bias among older children (H2). Consistent with our prediction, participants over-relied on testimony in a condition with high uncertainty and failed to use an optimal strategy to obtain reward. In Experiments 1 and 3, we also find support for the idea that testimony continues to influence choice-behavior beyond the preschool years. Though, when testimony is uncertain, impact is inversely related to age. Younger children appear to be especially sensitive to initial testimony and have relatively more difficulty updating their behavior. This finding stands in contrast to other research suggesting that younger children are particularly flexible in their ability to update beliefs in light of extant evidence (e.g., Gopnik et al., 2017; Lucas et al., 2014). The developmental differences observed here may represent a special case of interference in uncertain environments when testimony diverges from another source of, also uncertain, information.

Unpacking Strategy Use

One unexpected finding was that children who received unreliable and mixed reliability testimony or nonsocial cues did not transition to maximizing (i.e., exclusively choosing the highest rewarded option, the pattern of results observed in children in Plate et al. (2018), who were tested in the absence of an additional cue). Maximizing would be the optimal strategy in both the unreliable and mixed reliability conditions. Participants failed to maximize even though they were not following the cue's suggestions and were relying on the reward distribution (probability matching). However, previous research has also shown reduced rates of maximizing when the statistics of the task are made more difficult to distinguish (Plate et al., 2018).

Several explanations are plausible for participants' failure to maximize. First, perhaps participants simply did not have enough trials to capture maximizing in this task. Given that children typically probability match for a period of time prior to maximizing, it is possible that the addition of a second source of information required a prolonged period of probability matching prior to transitioning to maximizing. Second, it is possible that children considered alternative strategies beyond probability matching. One such strategy could have involved considering how "off" the suggestions were. Although there were not any deterministic trials (i.e., trials in which one could know the correct rock with 100% certainty based on the suggestion), it is possible that participants were trying to discern such a pattern. Future research, that includes trial types that systematically vary on determinism, is needed to better understand children's approaches.

Another feature that could have influenced strategy use was that the social learning context was ambiguous: children did not receive explicit information about the social agent's intentions, past experience, competence, or knowledge state. As these characteristics influence bias to trust in testimony (e.g., Buchsbaum et al., 2011; Corriveau et al., 2009; Einav & Robinson, 2011; Koenig & Harris, 2005; 2007; Kushnir et al., 2008; Pasquini et al., 2007), it would be useful to better understand how uncertainty affects the extent to which children rely on such information. Examining individual strategy approaches (see the online supplemental material) can provide some insight into how children interpreted the social manipulation regarding their own goals in the task. Finally, individual differences also have the potential to affect strategy use resulting from the learner's testimony bias (e.g., inhibitory control is associated with more bias to trust in testimony, Jaswal et al., 2014; see also Gilbert, 1991), attention to the underlying statistics (e.g., Browning et al., 2015), or both. Research related to individual differences would enhance our understanding of the complex interplay between the learner and the environment.

Limitations and Future Directions

We observed both similarities and differences in comparing the testimony and nonsocial cues. Differences between social and nonsocial conditions can be particularly difficult to interpret because of individuals' inclinations to attribute agency to nonsocial cues (e.g., Heider & Simmel, 1944; Wang et al., 2019). One possibility is that children privileged testimony over nonsocial information. This view is in line with research suggesting that social cues can direct attention and action above and beyond nonsocial cues (Tummeltshammer et al., 2014; Wu & Kirkham, 2010). Another possibility is that the differences were driven by the instructions. In other words, because participants were told that the nonsocial cue was incidental, perhaps they did not have any reason to think that the cue would help them find the reward. Despite this confound, participants in the reliable condition in both experiments used the cue. Yet, children could have been more willing to discount the nonsocial cue after having been told that it was not an intended aspect of the game.

There is a body of research on pointing gestures that suggests that pointing is a particularly strong cue for preschool age children and that these children may have a bias to search behind pointed locations (Palmquist et al., 2012; Palmquist & Jaswal, 2012). Evidence suggests that preschoolers are lured by pointers because they have difficulty overcoming a bias that these individuals are knowledgeable (as opposed to the alternate hypothesis that young children have difficulty inhibiting a response to the visual point display; Palmquist et al., 2018). Considering these issues, the testimony versus nonsocial contrast could cut both ways: Following the testimony in Experiment 1 could be too alluring because it is pointer-based and children are given instructions that might prime them to use the cue, therefore inflating the differences between Experiment 1 and Experiment 2. Alternatively, Experiment 2 could be unintentionally social because the cue is not entirely random, therefore diminishing the differences between Experiment 1 and Experiment 2. Related to these distinctions is the broader consideration of what "counts" as testimony. Much of the testimony research has focused on verbal testimony (Sobel & Finiasz, 2020); however, children receive input from others via myriad avenues (e.g., on social media; Antheunis et al., 2013; Freberg et al., 2011; Kaplan & Haenlein, 2010), in textbooks (Rey, 2012), and through nonverbal signaling (e.g., a parent simply pointing at a toy for the child to put away). Features that covary with the way input is conveyed could be meaningful for understanding when, why, and for how long children defer to others. For example, testimony provided in person could exert additional social pressure (e.g., to conform or follow a social norm), or it might increase the nonepistemic cues (e.g., the informant's social identity, position of power) that intensify deference to testimony (Harris et al., 2018; Jaswal & Kondrad, 2016). It is possible that having the confederate in the room with the child and either pointing (similar to Jaswal, 2010) or providing verbal testimony would have further amplified the bias to select the suggested location both in comparison to the nonsocial condition and across development.

In contrast to instances of verbal testimony, one feature of the current study is that the suggestion remains visible both during choice selection and during the presentation of outcomes. That the testimony cue remains visible (but without an in-person informant) may present a unique opportunity to treat the testimony as a single instance of data that can be compared against other data sources (Sobel & Kushnir, 2013). The complex nature of cues is precisely why investigations are needed to make headway in contrasting the mechanisms that underlie social learning and those that underlie nonsocial learning. We believe that the present research highlights an early step in this area of research.

Conclusion

Overall, this research provides evidence regarding lingering testimony bias across childhood. Children were sensitive to the reward structure and the statistics of the testimony. However, children had difficulty adjusting behavior when testimony was uncertain, and younger children were disproportionately likely to rely on testimony under uncertainty. Younger children were also more strongly influenced by initial testimony prior to an unexpected change, an effect that decreased with age. Through these experiments, we shed light on conditions that can strengthen bias to defer to testimony even when children are at ages in which they otherwise demonstrate rational behavior.

References

Antheunis, M. L., Tates, K., & Nieboer, T. E. (2013). Patients' and health professionals' use of social media in health care: Motives, barriers and expectations. *Patient Education and Counseling*, *92*(3), 426–431. https://doi.org/10.1016/j.pec.2013.06.020

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men* (pp. 177–190). Carnegie.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Scheipl, F. (2012). *R package 'lme4'*. R Foundation for Statistical Computing. https://www.R-project.org/

Birch, S. A. J., Severson, R. L., & Baimel, A. (2020). Children's understanding of when a person's confidence and hesitancy is a cue to their credibility. *PLoS ONE*, *15*(1), e0227026. https://doi.org/10.1371/journal.pone.0227026

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology*, *42*(3), 500–513. https://doi.org/10.1037/0012-1649.42.3.500

Bridgers, S., Buchsbaum, D., Seiver, E., Griffiths, T. L., & Gopnik, A. (2016). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology*, *52*(1), 9–18. https://doi.org/10.1037/a0039830

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. https://doi.org/10.1038/nn.3961

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, *120*(3), 331–340. https://doi.org/10.1016/j.cognition.2010.12.001

Corriveau, K. H., Meints, K., & Harris, P. L. (2009). Early tracking of informant accuracy and inaccuracy. *British Journal of Developmental Psychology*, *27*(2), 331–342. https://doi.org/10.1348/026151008X310229

Einav, S., & Robinson, E. J. (2011). When being right is not enough: Four-year-olds distinguish knowledgeable informants from merely accurate informants. *Psychological Science*, *22*(10), 1250–1253. https://doi.org/10.1177/0956797611416998

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Freberg, K., Graham, K., McGaughey, K., & Freberg, L. A. (2011). Who are the social media influencers? A study of public perceptions of personality. *Public Relations Review*, *37*(1), 90–92. https://doi.org/10.1016/j.pubrev.2010.11.001

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107–119. https://doi.org/10.1037/0003-066X.46.2.107

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., & Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), 7892–7899. https://doi.org/10.1073/pnas.1700811114

Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Gualtieri, S., Buchsbaum, D., & Denison, S. (2020). Exploring information use in children's decision-making: Base-rate neglect and trust in testimony. *Journal of Experimental Psychology: General*, *149*(8), 1527–1536. https://doi.org/10.1037/xge0000726

Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition*, *132*(3), 335–341. https://doi.org/10.1016/j.cognition.2014.04.013

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Harvard University Press. https://doi.org/10.4159/harvard.9780674065192

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*(3), 505–524. https://doi.org/10.1111/j.1467-8624.2006.00886.x

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*, 251–273. https://doi.org/10.1146/annurev-psych-122216-011710

Haun, D. B., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development*, *82*(6), 1759–1767. https://doi.org/10.1111/j.1467-8624.2011.01666.x

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243–259. https://doi.org/10.2307/1416950

Hermansen, T. K., Ronfard, S., Harris, P. L., Pons, F., & Zambrana, I. M. (2021). Young children update their trust in an informant's claim when experience tells them otherwise. *Journal of Experimental Child Psychology*, *205*, 105063. https://doi.org/10.1016/j.jecp.2020.105063

Hermes, J., Behne, T., Bich, A. E., Thielert, C., & Rakoczy, H. (2018). Children's selective trust decisions: Rational competence and limiting performance factors. *Developmental Science*, *21*(2), e12527. https://doi.org/10.1111/desc.12527

Hermes, J., Rakoczy, H., & Behne, T. (2019). Making sense of conflicting information: A touchscreen paradigm to measure young children's selective trust. *Infant and Child Development*, *28*(1), e2119. https://doi.org/10.1002/icd.2119

Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive Psychology*, *61*(3), 248–272. https://doi.org/10.1016/j.cogpsych.2010.06.002

Jaswal, V. K., & Kondrad, R. L. (2016). Why children are not always epistemically vigilant: Cognitive limits and social considerations. *Child Development Perspectives*, *10*(4), 240–244. https://doi.org/10.1111/cdep.12187

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, *21*(10), 1541–1547. https://doi.org/10.1177/0956797610383438

Jaswal, V. K., Pérez-Edgar, K., Kondrad, R. L., Palmquist, C. M., Cole, C. A., & Cole, C. E. (2014). Can't stop believing: Inhibitory control and resistance to misleading testimony. *Developmental Science*, *17*(6), 965–976. https://doi.org/10.1111/desc.12187

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277. https://doi.org/10.1111/j.1467-8624.2005.00849.x

Koenig, M. A., & Harris, P. L. (2007). The basis of epistemic trust: Reliable testimony or reliable sources? *Episteme*, *4*(3), 264–284. https://doi.org/10.3366/E1742360007000081

Kushnir, T., Wellman, H. M., & Gelman, S. A. (2008). The role of preschoolers' social understanding in evaluating the informativeness of causal interventions. *Cognition*, *107*(3), 1084–1092. https://doi.org/10.1016/j.cognition.2007.10.004

Li, X., & Yow, W. Q. (2018). Willingness to revise own testimony: 3- and 4-year-olds' selective trust in unexpected testimony from accurate and inaccurate informants. *Journal of Experimental Child Psychology*, *173*, 1–15. https://doi.org/10.1016/j.jecp.2018.03.008

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*(2), 284–299. https://doi.org/10.1016/j.cognition.2013.12.010

Magner, U. I., Schwonke, R., Aleven, V., Popescu, O., & Renkl, A. (2014). Triggering situational interest by decorative illustrations both fosters and hinders learning in computer-based learning environments. *Learning and Instruction*, *29*, 141–152. https://doi.org/10.1016/j.learninstruc.2012.07.002

Mcloughlin, N., Finiasz, Z., Sobel, D. M., & Corriveau, K. H. (2021). Children's developing capacity to calibrate the verbal testimony of others with observed evidence when inferring causal relations. *Journal of Experimental Child Psychology*, *210*, 105183. https://doi.org/10.31219/osf.io/zuqwm

Palmquist, C. M., & Jaswal, V. K. (2012). Preschoolers expect pointers (even ignorant ones) to be knowledgeable. *Psychological Science*, *23*(3), 230–231. https://doi.org/10.1177/0956797611427043

Palmquist, C. M., Burns, H. E., & Jaswal, V. K. (2012). Pointing disrupts preschoolers' ability to discriminate between knowledgeable and ignorant informants. *Cognitive Development*, *27*(1), 54–63. https://doi.org/10.1016/j.cogdev.2011.07.002

Palmquist, C. M., Kondrad, R. L., & Norris, M. N. (2018). Follow my point? Preschoolers' expectations of veridicality disrupt their understanding of deceptive points. *Cognitive Development*, *48*, 190–202. https://doi.org/10.1016/j.cogdev.2018.08.009

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*(5), 1216–1226. https://doi.org/10.1037/0012-1649.43.5.1216

Plate, R. C., Fulvio, J. M., Shutts, K., Green, C. S., & Pollak, S. D. (2018). Probability learning: Changes in behavior across time and development. *Child Development*, *89*(1), 205–218. https://doi.org/10.1111/cdev.12718

R Core Team. (2019). R: A language and environment for statistical computing (Version 3.1.2). R Foundation for Statistical Computing.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*(3), 216–237. https://doi.org/10.1016/j.edurev.2012.05.003

Riggs, A. E., Alibali, M. W., & Kalish, C. W. (2015). Leave her out of it: Person-presentation of strategies is harmful for transfer. *Cognitive Science*, *39*(8), 1965–1978. https://doi.org/10.1111/cogs.12224

Riggs, A. E., Alibali, M. W., & Kalish, C. W. (2017). Does it matter how Molly does it? Person-presentation of strategies and transfer in mathematics. *Contemporary Educational Psychology*, *51*, 315–320. https://doi.org/10.1016/j.cedpsych.2017.09.001

Ronfard, S., & Lane, J. D. (2018). Preschoolers continually adjust their epistemic trust based on an informant's ongoing accuracy. *Child Development*, *89*(2), 414–429. https://doi.org/10.1111/cdev.12720

Ronfard, S., Lane, J. D., Wang, M., & Harris, P. L. (2017). The impact of counter-perceptual testimony on children's categorization after a delay. *Journal of Experimental Child Psychology*, *163*, 151–158. https://doi.org/10.1016/j.jecp.2017.06.006

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, *15*(3), 436–447. https://doi.org/10.1111/j.1467-7687.2012.01135.x

Sobel, D. M., & Finiasz, Z. (2020). How children learn from others: An analysis of selective word learning. *Child Development*, *91*(6), e1134–e1161. https://doi.org/10.1111/cdev.13415

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, *120*(4), 779–797. https://doi.org/10.1037/a0034191

Sobel, D. M., Buchanan, D. W., Butterfield, J., & Jenkins, O. C. (2010). Interactions between causal models, theories, and social cognitive development. *Neural Networks*, *23*(8–9), 1060–1071. https://doi.org/10.1016/j.neunet.2010.06.004

Starling, S. J., Reeder, P. A., & Aslin, R. N. (2018). Probability learning in an uncertain world: How children adjust to changing contingencies. *Cognitive Development*, *48*, 105–116. https://doi.org/10.1016/j.cogdev.2018.06.008

Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants track the reliability of potential informants. *Psychological Science*, *25*(9), 1730–1738. https://doi.org/10.1177/0956797614540178

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360. https://doi.org/10.1146/annurev.psych.59.103006.093707

Wang, F., Tong, Y., & Danovitch, J. (2019). Who do I believe? Children's epistemic trust in internet, teacher, and peer informants. *Cognitive Development*, *50*, 248–260. https://doi.org/10.1016/j.cogdev.2019.05.006

Wickham, H. (2016). *ggplot2-Elegant Graphics for Data Analysis*. Springer International Publishing.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*(2), 118–136. https://doi.org/10.1016/j.jecp.2010.04.014

Received January 22, 2021
Revision received June 18, 2021
Accepted July 16, 2021 ■