Accuracy, Confidence, and Calibration: How Young Children and Adults Assess Credibility

Elizabeth R. Tenney, Jenna E. Small, Robyn L. Kondrad, Vikram K. Jaswal, and Barbara A. Spellman University of Virginia

Do children and adults use the same cues to judge whether someone is a reliable source of information? In 4 experiments, we investigated whether children (ages 5 and 6) and adults used information regarding accuracy, confidence, and calibration (i.e., how well an informant's confidence predicts the likelihood of being correct) to judge informants' credibility. We found that both children and adults used information about confidence and accuracy to judge credibility; however, only adults used information about informants' calibration. Adults discredited informants who exhibited poor calibration, but children did not. Requiring adult participants to complete a secondary task while evaluating informants' credibility impaired their ability to make use of calibration information. Thus, children and adults may differ in how they infer credibility because of the cognitive demands of using calibration.

Keywords: credibility, confidence, accuracy, calibration, testimony

People rely on others to learn much of what they know. They cannot see that the earth is round or instinctively know that Abraham Lincoln was 6 ft 4 in.—they learn these facts by relying on other people's testimony (e.g., Koenig, Clément, & Harris, 2004; Harris & Richert, 2008). Thus, to ensure the truth of what they believe about the world, people may seek out reliable and credible informants. But how do people determine the credibility of their informants, and do their strategies change from childhood to adulthood?

Confidence and Credibility

One prevalent cue that adults use to judge someone's credibility is confidence (e.g., Brewer & Burke, 2002; Penrod & Cutler, 1995; Sporer, Penrod, Read, & Cutler, 1995). Some researchers have even surmised that whether a person exhibits confidence is the most important determinant of whether that person will be believed. In a study that used a mock-trial paradigm, researchers manipulated 10 variables that were likely to affect a witness's credibility, including the witness's confidence (80% vs. 100%)

This article was published Online First March 28, 2011.

Elizabeth R. Tenney, Jenna E. Small, Robyn L. Kondrad, Vikram K. Jaswal, and Barbara A. Spellman, Department of Psychology, University of Virginia.

We thank Le-Anh Bui, Mayra Perez, Alison Setia, Carrington Croft, Debby Chen, Chris Holland, Jessica Bogacik, and Zane Shuck for their assistance in data collection. We thank Mieke VanderBorght and Monica Erbacher for thoughtful discussions, and the two children who posed for pictures for the storybook. We also extend our thanks to the actors who played the witnesses in Experiment 4 and especially thank Molly Bishop Shadel for her expertise in directing and filming the actors.

Correspondence concerning this article should be addressed to Elizabeth R. Tenney, Department of Psychology, University of Virginia, 102 Gilmer Hall, P.O. Box 400400, Charlottesville, VA 22904-4400. E-mail: tenney@virginia.edu

confident in an identification), integrity of a lineup, presence of a weapon at the crime, and conditions of identification (e.g., whether the culprit was wearing a disguise; whether the identification occurred 2 days or 2 weeks after the crime). Although past research had suggested that a number of variables affect jurors' decisions, the witness's confidence was the only variable to reliably and significantly increase belief in the defendant's culpability (Cutler, Penrod, & Stuve, 1988).

Given how compelling confidence seems to be in establishing credibility, some researchers have suggested that there might be a confidence heuristic—meaning that people use statements of confidence as a shortcut when deciding the worth of informants or advisors (Price & Stone, 2004; Thomas & McFadyen, 1995). Others have explained why confidence is important, at least to adults, using the presumption of calibration hypothesis, which posits that barring other information, people initially assume informants have some insight into the quality of their own knowledge, and as a result, people assume an informant's high confidence is a good indicator of high accuracy. However, new details about how a particular informant uses confidence terms (e.g., revealing over- or underconfidence) will dispel that assumption (Tenney & Spellman, in press; Tenney, Spellman, & MacCoun, 2008).

There is reason to believe that young children use some of the same cues as adults, including the degree of informant confidence, when evaluating credibility. Anecdotally, parents and caretakers of young children learn that statements such as "You might want to take a jacket; you could be cold later" are less effective than "You need to take a jacket; you will be cold later." Research has confirmed that children are more likely to believe a confident informant over an uncertain one (e.g., Harris, 2007; Jaswal & Malone, 2007; Sabbagh & Baldwin, 2001; Sabbagh, Wdowiak, & Ottaway, 2003). Four-year-old children, for example, were more likely to search for an object in a box that a puppet identified by saying, "I know it's in the red box," than in a box that another puppet identified by saying, "I think it's in the blue box" or "I

guess it's in the blue box" (Moore, Bryant, & Furrow, 1989). Three-year-olds were also more likely to believe that a spoon-like object was a key when a speaker asserted with confidence, "This is a key" than when the speaker said, "I think this is a key" (Jaswal & Malone, 2007). Together, the evidence suggests that children are sensitive to variations in confidence when determining whether they should listen to someone.

Accuracy and Credibility

In addition to confidence, when possible, people use an informant's history of accuracy to determine the informant's credibility. Research with adults has shown that informants who have been inconsistent or otherwise made a factual error—even about peripheral details—seem less trustworthy about other things they say (Borckardt, Sprohge, & Nash, 2003).

There is also evidence that children prefer informants who have been accurate rather than inaccurate in the past. A good deal of recent research has shown that preschoolers are more likely to seek help and use information from an informant who has been accurate in the past compared with one who has been inaccurate (usually on the basis of at least three instances of accuracy or inaccuracy). For example, they prefer labels for novel objects offered by the formerly accurate informant over labels offered by the formerly inaccurate informant (e.g., Jaswal, McKercher, & VanderBorght, 2008; Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005; Pasquini, Corriveau, Koenig, & Harris, 2007). Additionally, although children typically prefer to learn new words from adults rather than peers (Taylor, Cartwright, & Bowden, 1991), if a particular adult has previously been unreliable (e.g., labeling a shoe as a "glass" and making similar mistakes in four trials), young children instead endorse new vocabulary provided by a reliable child informant (Jaswal & Neely, 2006). Thus, a history of accuracy (on the basis of previous instances of proving to be correct vs. incorrect) is an important determinant of credibility.

Confidence Plus Accuracy Reveals Calibration

What happens when people have information about both an informant's confidence and accuracy? In that case, people can make inferences about how well-calibrated he or she is—that is, how well that person's confidence matches his or her likelihood of being correct. Informants who are well calibrated have high confidence when they are likely to be right and low confidence when they have reason to believe they are wrong. In contrast, informants who are poorly calibrated have confidence that is independent of their likelihood of being correct.

Previous research has shown that adults are sensitive to information about informants' calibration (Tenney, MacCoun, Spellman, & Hastie, 2007; Tenney et al., 2008). In these studies, participants read about two witnesses to a car accident. The high-confidence witness was sure not only about what he saw during the accident but also about two peripheral events (e.g., that he had a meeting at work that day and that he took his dog to the vet that day). The cautious witness was also sure about what he saw during the accident and about one peripheral event, but he was not sure about the other peripheral event. At this point, participants rated the high-confidence witness as more credible than the cautious witness and also tended to believe his version of the accident.

Subsequently, participants learned that a private investigator had discovered that the high-confidence witness was wrong about one of the peripheral events (about which he had been highly confident) and that the cautious witness was wrong about the peripheral event about which he had been uncertain. When given a second opportunity to rate the credibility of the witnesses and to indicate whose account of the accident they believed, participants switched to favor the cautious but well-calibrated witness. (In another study, in which the cautious witness was poorly calibrated, both witnesses' credibility suffered, and participants continued to prefer the high-confidence witness.)

Thus, it is not the case that adults find the most confident informants to be always the most credible. Instead, there is an interaction between confidence and accuracy, such that the amount of damage an error causes to credibility depends on how confident the informant had been when making the error. When informants make errors (e.g., misremember what day a meeting occurred), those who expressed less confidence in the incorrect information were seen more positively. Adults were able to remember the degree of confidence two informants expressed in three assertions, pair the level of confidence with accuracy for two of the assertions, make inferences about the informants' individual confidence—accuracy relations, and use those inferences to assess the veracity of the informants' remaining assertions.

Overview of Current Experiments

In the present studies, we investigate the extent to which children integrate confidence and accuracy information to evaluate how well calibrated an informant is, and we investigate some possible explanations for why they do not. (See Table 1 for experimental designs.)

Experiment 1a compared children's and adults' use of calibration information. We designed a storybook with a story analogous to the car accident scenario (described above) that would be of interest to children and tolerable for adults. The question was whether children, like adults, would prefer a cautious, well-calibrated informant to an overly confident one. To foreshadow our findings, adults preferred the well-calibrated witness, but children preferred the overly confident one.

Experiment 1b replicated the findings with children in Experiment 1a with the use of a more typical paradigm for investigating children's reliance on informant testimony. In Experiment 1b, two informants provided labels for familiar objects on a video with varying levels of accuracy and confidence, and we measured which informants' labels children adopted for novel objects. As in Experiment 1a, children relied on an overconfident informant rather than a well-calibrated one.

Experiments 2–4 were designed to investigate why children did not seem to make use of calibration information. Specifically, we were interested in whether children had the requisite understanding of the components of calibration. In Experiment 2, we held the

¹ We use calibration as a general covering term to refer to different types of confidence–accuracy correspondence. More precise terms for the within-person confidence–accuracy correlations we aim to capture in our experimental manipulations are *monitoring resolution* or *discrimination accuracy* (see discussion of these terms in Koriat, 1997, and Yaniv, Yates, & Smith, 1991).

Table 1
Experimental Conditions

Experiment and manipulation	Description of informant	Informant confidence	Informant accuracy
Experiments 1 & 4 (calibration)	Confident (poorly-calibrated)	Confident re: A	Correct
		Confident re: B	Incorrect
	Cautious (well-calibrated)	Confident re: A	Correct
		Hesitant re: C	Incorrect
Experiment 2 (confidence)	Confident	Confident re: A	Correct
		Confident re: B	Correct
	Cautious	Confident re: A	Correct
		Hesitant re: B	Correct
Experiment 3 (accuracy)	Accurate	Confident re: A	Correct
		Confident re: B	Correct
	Inaccurate	Confident re: A	Correct
		Confident re: C	Incorrect

Note. A, B, and C refer to noncritical (i.e., collateral) statements made by the informants. B and C were counterbalanced across informants in Experiments 1, 3, and 4.

accuracy of two informants constant while varying their confidence. In contrast, in Experiment 3 we held informants' confidence constant while varying their accuracy. Finally, in Experiment 4, we explored whether adults would keep track of confidence and accuracy but make less use of calibration information when their cognitive resources were depleted.

Experiment 1a: Children's and Adults' Use of Calibration

In Experiment 1a, 5- to 6-year-old children and college-age adults saw a picture book titled *The Mystery of the Broken Window*. In the story, two girls were witnesses to an accident in which someone hit a baseball through a window, but each witness claimed that a different boy was at fault. The materials were analogous in structure to the car accident story materials used in previous research (Tenney et al., 2007, 2008). Thus, we expected the adults to rely on calibration; the question was whether children would also do so.

Method

Materials and design. Participants heard testimony about an accident from two witnesses, and they were twice asked to decide which witness they believed—once before certain facts about the accident were revealed and once after. An experimenter read a storybook to each participant in which two boys were playing at recess and one of the boys hit a ball through a window. To figure out who was responsible, a teacher asked two witnesses who saw the incident what happened.

The two witnesses each remembered two details of the incident but differed in how confident they were. The confident witness confidently described two details about the accident, and then confidently accused one boy of being the perpetrator: "I know it was sunny, I know the ball was red, and I know that Tyler hit the ball through the window." The cautious witness confidently described one detail about the accident, was cautious about another, and then confidently accused a different boy of being the perpetrator: "I know it was sunny; I think maybe the ball was blue, but I'm not sure; and I know that Kenny hit the ball through the

window." Crucially, the two witnesses were both confident in their accusation of who hit the ball through the window. Pictures of each object or person the witnesses mentioned accompanied their testimony, and pictures of the two witnesses looking confident (or uncertain) underscored their confidence level about each fact (see Figure A1 in the Appendix). The experimenter repeated the contradictory accusations of who was responsible while pointing to the pictures, and participants were asked which witness they believed (Time 1 question: "Emily said, 'I know Tyler hit the ball right through the window,' and Jessica said, 'I know Kenny hit the ball right through the window.' Whom do you believe, Emily or Jessica?").

Next, participants learned the veracity of some aspects of the witnesses' stories. Using pictures, the experimenter reminded children what each witness had said and explained that both witnesses had been correct about the weather (it was indeed sunny) and incorrect about the color of the ball (it was actually white, not red or blue). Given this information, the confident witness was poorly calibrated (because she was confident regardless of whether she was right), and the cautious witness was well calibrated (because she was confident when right, cautious when wrong). The accusations about which boy broke the window were repeated, and children were again asked which witness they believed (Time 2). Finally, participants learned that Kenny had hit the ball through the window, and he apologized to the teacher. The order in which the witnesses testified and which girl was the confident witness was counterbalanced across participants.

Adult participants and procedure. Thirty-three undergraduates at a selective public university (M=19 years 10 months; range = 18-22 years; 15 women) participated in a 10-min session and received psychology course credit. Participants were tested individually in a laboratory and sat at a table next to an experimenter. The experimenter read a storybook to each participant and pointed to the objects or characters in the pictures when speaking about them. In addition, when appropriate in the story, the experimenter changed her voice to be confident or uncertain to emphasize what the informant said.

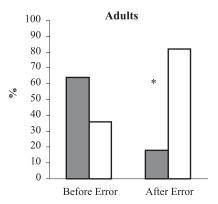
Child participants and procedure. Forty-nine 5- and 6-year-old children (M = 5 years 11 months; range = 5 years 0

months to 6 years 11 months; 24 girls) participated in a 10-min session. Participants were recruited from a database of community families and were from predominantly Caucasian, middle-class backgrounds. Just as for the adults, the children were tested individually and sat at a table next to an experimenter. The children's parents were given the option of being in the room (out of the child's direct line of vision), but most parents decided to remain in a waiting area. Just as for the adults, the experimenter read a storybook to each child with appropriate pointing and voice changing.

Results and Discussion

At Time 1, both adults and children preferred the confident witness; but at Time 2, after the errors were revealed, children continued to prefer the confident witness, whereas most adults (as in previous research) switched to the cautious (well-calibrated) witness (see Figure 1.)

Adult results. There were no significant effects of order or actor, so these variables were not considered further. As the top panel of Figure 1 shows, at Time 1, before errors were revealed, 21 adult participants (64%) believed the confident witness and 12 (36%) believed the cautious witness. At Time 2, after errors were revealed, the pattern of responses switched; now 27 participants



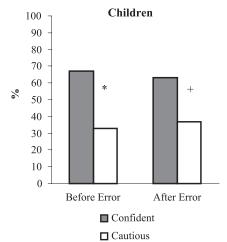


Figure 1. Percentage of participants who chose to believe the confident and cautious witnesses before and after errors were revealed in Experiment 1a. $^{\dagger}p < .10. ^*p < .05$.

(82%) believed the cautious witness and only six (18%) believed the confident witness. A 1×4 chi-square goodness-of-fit test² comparing all possible combinations of choices across Time 1 and Time 2 revealed that the changes in choices were not random, $\chi^2(3, N=33)=16.09, p=.001$. Fifteen participants (45%) switched from the confident to the cautious witness, which was significantly greater than would be expected by chance (i.e., 8.3 participants [25%]). No participants switched in the opposite direction. Six participants (18%) chose the confident witness twice, and 12 (36%) chose the cautious witness twice, which would be expected by chance. These results replicate previous research with new materials (Tenney et al., 2007, 2008). Adults preferred informants who recognized the limits of their own knowledge and who expressed uncertainty where appropriate.

Child results. Unlike adults, children did not change which witness they believed after learning about the errors. There were no significant effects of participant age (5 vs. 6) or whether the confident or cautious witness testified first; however, participants did show a preference for one of the two actors regardless of whether that actor was confident. We have no ready explanation for this; in the subsequent studies with the same actors, no preference was noted, so we assume that this was a spurious effect. Collapsing across order and actor, as the lower panel in Figure 1 shows, at Time 1, before errors were revealed, 33 child participants (67%) believed the confident witness and 16 (33%) believed the cautious witness; this preference was significantly different from what would be expected by chance, $\chi^2(1, N = 49) = 5.90$, p = .015.

It is worth emphasizing that the witnesses differed in confidence on just one detail—the color of the ball. They were both confident about the weather and also about the central issue—who had broken the window. In previous research where children have been shown to be sensitive to speaker certainty (e.g., Jaswal & Malone, 2007; Moore, Bryant, & Furrow, 1989; Sabbagh & Baldwin, 2001), children evaluated the credibility of the actual statement about which the speaker had been hesitant or confident. In contrast, in the current study, children used a difference in the confidence of the two witnesses about a rather trivial detail as the basis for deciding whose confidently asserted accusation was more credible.

At Time 2, after errors were revealed and in contrast to the results from adult participants, the pattern of responses was consistent with Time 1: Thirty-one children (63%) believed the confident witness, and 18 (37%) believed the cautious witness, $\chi^2(1, N=49)=3.44, p=.063$. A 1×4 chi-square goodness-of-fit test comparing all possible combinations of choices across Time 1 and Time 2 was significant, $\chi^2(3, N=49)=31.57, p<.001$. Twenty-eight children (57%) chose the confident witness at both times, which was significantly greater than would be expected by chance (i.e., 12.3 children [25%]). Five (10%) switched from

² We used a goodness-of-fit test because of the nonindependent nature of the within-subject design. We considered using a sign change test, but then a finding of nonsignificance could mean two different things: (a) Participants were not changing their Time 1 to Time 2 preference, or (b) participants were changing in various directions and canceling each other out. A goodness-of-fit test reveals whether participants were significantly changing their preferences from Time 1 to Time 2 and in which direction.

confident to cautious, and three (6%) switched from cautious to confident—both of which were significantly fewer than expected by chance. Thirteen (25.5%) chose the cautious witness twice, which was not different from chance. Thus, unlike adults, children did not have a preference for well-calibrated informants; rather, they trusted informants who were the most confident, regardless of whether that confidence was warranted.

Experiment 1b: Children's Use of Calibration in an Object-Naming Task

Experiment 1a suggests that children monitor and rely on confidence rather than calibration to judge informant credibility. One limitation of this experiment, however, is that it uses only one paradigm (a storybook) to test this prediction. Perhaps there was something about our materials, such as the complexity of the story or the demands on memory, that made it especially likely that children would rely on confidence. We explored this possibility in Experiment 1b by using a more typical paradigm for investigating children's appraisal of informants' testimony—namely, a wordlearning scenario (e.g., Koenig et al., 2004). Two informants provided labels for familiar and novel objects with varying levels of confidence and accuracy. This procedure allowed children to evaluate the veracity of what an informant claimed (e.g., that a toy tree was "a glass") using their own knowledge rather than having to wait to learn from the experimenter what was true, as in Experiment 1a. As in Experiment 1a, one informant was confident but poorly calibrated, whereas the other was less confident but well calibrated.

Method

Participants. Twenty-six 5- and 6-year-olds (M = 5 years 11 months; 14 girls) participated in a single 10-min session. Children were recruited from the same database of community families as in Experiment 1a, but they had not participated in that experiment.

Materials and design. We created a short video with two female actors who played the part of confident and cautious labelers. The actors sat side by side at a table and wore very different shirts to help children distinguish them (a blue buttondown shirt vs. a bright red sweater). At the start of each trial, the object that the actors were going to label was in the middle of the table. The video consisted of two induction trials using objects that were familiar (a basketball and a plastic toy tree) and one test trial with an object that was novel (a scrubber with a long handle). The purpose of the induction trials was to show that one actor was confident but poorly calibrated in the names she gave objects (i.e., confident both when right and when wrong), and the other actor was more cautious overall but well calibrated (confident when right, cautious when wrong). At the test trial, both actors were confident in the different names they provided for the novel object.

Procedure. To begin the session, the experimenter introduced the game and the two actors while showing a still frame from the video. The experimenter said,

Do you see the two people on the screen? You know what? Those are two of my friends! Yeah, this is my friend Liz (point), and this is my friend Jane (point). My friends have all kinds of fun things to show you today! They are going to show you some things up here (point to

computer screen) and tell you what those things are called. Does that sound like fun? Let's see what they have.

At Induction Trial 1, both actors were confident and correct. One actor pointed to the object in front of her and said, "Look! I know this is a basketball. Yup, it's definitely a basketball." The other actor did the same but called the object a "ball." Then the experimenter paused the video and said, "So she said she knew it was called a basketball, and she said she knew it was called a ball. Do you think it's a basketball or a ball?" (Labels were repeated in the order the actors had given them.) Unless the child spontaneously answered that both names were correct, the experimenter asked if the object could also be called by the other name. Then, to emphasize that both actors had been correct, the experimenters asked, "Do you think one of my friends was saying something wrong? Or was neither of my friends wrong?" After correcting the child if necessary, the experimenter said, "That's right, so they were both right!"

At Induction Trial 2, both actors provided incorrect labels for a familiar object; however, one actor was confident and the other was cautious. In the video, the confident actor pointed to a plastic toy tree and said, "Look! I know that's a telephone. Yup, it's definitely a telephone." The cautious actor pointed to the same tree and said, "Look! Hmm, I don't know. I think maybe it's a glass. Maybe it's a glass?" and she shrugged her shoulders. After the actors spoke, the experimenter reminded the children what was said: "So she said she knew it was called a telephone, and she said she thought maybe it was called a glass. Do you think it's a telephone or a glass?" After children responded, the experimenter asked, "Do you think one of my friends was saying something wrong? Or were both of my friends wrong?" Next, the experimenter reminded children again of what each actor had said: "So, they were both wrong, right? But she said she knew it was a telephone, and she said she thought maybe it was a glass."

At the test trial, both actors provided different labels for a novel object (a scrubber with a long handle) with high confidence. One said, "Look! I know this is a jeter. Yup, it's definitely a jeter," and the other said exactly the same thing but referred to the object as a "dax." Then experimenters reiterated what the actors had said and asked, "Do you think it's a jeter or a dax?"

The order that the confident and cautious actors spoke, the actors who played each part, and the labels the actors provided for objects were counterbalanced across participants.

Results and Discussion

Children in Experiment 1b, like those in Experiment 1a, preferred the confident rather than the well-calibrated witness. There were no effects of age or sex of participant so those variables were not explored further.

At Induction Trial 1, as expected, children were essentially evenly split between the two actors (who had both been confident and correct): Ten chose the label from the actor who later ended up being cautious, 13 chose the label from the actor who ended up being always confident, and three said, without prompting, that both labels were right. There was an item effect such that the majority of children believed that the actor who labeled the object a basketball rather than a ball was right (i.e., 22 out of the 23 children who did not spontaneously choose both labels). When

asked whether the object could also be called a ball [basketball], all children agreed except for one.

At Induction Trial 2, when asked directly whether one or both actors said anything wrong, all children said both actors were wrong except for three children who believed that the confident actor's label was correct. The four children who failed the manipulation check at either induction trial were excluded from the test trial results reported below (but their exclusion did not affect direction or significance of effects).

At the test trial, when both actors provided labels for novel objects with equally high confidence, 17 children (77%) chose the label provided by the actor who had always been confident rather than the actor who had been cautious once but well-calibrated, $\chi^2(1, N=22)=6.55, p=.01$. These results provide further evidence that children rely on high confidence more than calibration to determine the credibility of informants and show that children's failure to use calibration information in Experiment 1a was not because of the particular materials used in that study.

Why don't children seem to make use of calibration information? One possibility is that they do not notice or extract the relevant confidence and accuracy information from the story (before it can even be combined into calibration information). The next two studies were designed to investigate this possibility.

Experiment 2: Children's Use of Confidence

Experiment 2 was designed to investigate specifically whether children would prefer an informant who had been confident in the past over one who had been uncertain. We revised the storybook from Experiment 1a so that both witnesses were equally accurate (right about both details), but one was confident and the other was cautious. We also added a manipulation check at the end of the experiment to determine whether children could identify the confident and cautious witnesses. On the basis of previous research (e.g., Jaswal & Malone, 2007; Moore et al., 1989) and our own findings with children from Experiments 1a and 1b, we expected that children would successfully keep track of and use confidence when deciding whom to believe.

Method

Participants and procedure. Twenty-four 5- and 6-year-olds (M = 5 years 11 months; 12 girls) participated in a single 10-min session. Children were recruited from the same database of community families as in Experiments 1a and 1b, but they had not participated in those experiments. The procedure was the same as in Experiment 1a.

Materials and design. The story and storybook were similar to Experiment 1a. The story up to the Time 1 question was the same as in Experiment 1a except that both the confident and cautious witnesses claimed the ball was white (rather than one saying it was red and the other saying it was blue). Time 2 was similar to Experiment 1a except that instead of learning that both witnesses had made an error, children learned that both witnesses had been correct about the two details (the sunny weather and the white ball). Thus, the witnesses varied on confidence but each was correct about both details. As in Experiment 1a, children chose which witness they believed at Time 1 and again at Time 2. Afterward, participants answered yes or no to manipulation check

questions about how confident each witness had been (e.g., "Was Emily always sure?"). The order in which the two witnesses testified and which girl was the confident witness were counterbalanced across participants.

Results and Discussion

Preliminary analyses showed no significant differences of age (5 vs. 6), sex of participant, order, or actor, so we collapsed across these variables. Seventeen participants out of 24 (71%) passed the manipulation check questions indicating they recognized the difference between the confidence levels of the witnesses, and this effect was significantly different from chance, $\chi^2(1, N = 24) = 4.17$, p = .042.

As Figure 2 shows, the children were sensitive to the witnesses' past confidence. At Time 1, of the participants who passed the manipulation check, 13 participants (76.5%) preferred the confident witness and four (23.5%) preferred the cautious witness; this preference for the confident witness was statistically significant, $\chi^2(1, N=17)=4.77, p=.029$. Thus, most children seemed to be sensitive to varying confidence levels and preferred to side with the most confident witness.

At Time 2, after learning that both witnesses were accurate about the details of the case, 10 (59%) of the participants who passed the manipulation checks preferred the confident witness and seven (41%) preferred the cautious one, which was not a significant difference $\chi^2(1, N=17)=0.53, p=.467$. There were no actor effects. Thus, it seems that when children discovered that both witnesses had been accurate, differences in past levels of confidence no longer made a difference.

Experiment 3: Children's Use of Accuracy

Experiment 2 demonstrated that children used confidence to assess credibility before they had information about accuracy. In Experiment 3, we sought to determine whether children were sensitive to (relatively subtle) differences in the accuracy of informants. We used a revised storybook, in which both witnesses were equally confident about each fact, but one was correct about each one and the other made an error. As in Experiment 2, we also added a manipulation check at the end of the experiment, this time to determine whether children could remember which witness had been more accurate. In typical experiments with children that vary accuracy rates, the informants make repeated errors (e.g., Birch, Vauthier, & Bloom, 2009; Clément et al., 2004; Jaswal & Neely, 2006; Koenig & Harris, 2004; but see Pasquini et al., 2007, for the exception). In this experiment, we wanted to determine whether children were sensitive to a single difference in accuracy. If we found that children prefer the accurate informant using our storybook paradigm, then we could rule out the possibility that children do not use calibration information simply because they do not

³ We analyzed the results regarding which witness participants believed, with and without those who failed this manipulation check. The pattern of results and significance was the same in both cases at Time 2, but at Time 1, when those who failed the manipulation check were considered in analysis, 16 participants (67%) preferred the confident witness and eight (33%) preferred the cautious one. This preference was in the expected direction but not significant, $\chi^2(1, N = 24) = 2.67, p = .102$.

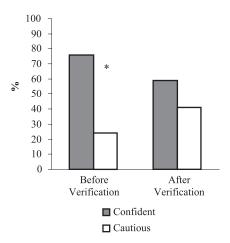


Figure 2. Percentage of participants (who passed the manipulation check) who selected the confident and cautious witnesses before and after learning that each had been right about the details they provided in Experiment 2. *p < .05.

notice the errors or do not consider a single error about a relatively minor detail (e.g., the color of the ball that went through a window) to be important.

Method

Participants and procedure. Twenty-four 5- and 6-year-olds (M = 5 years 10 months; 12 girls) participated in a single 10-min session. Children were recruited from the same database of community families, but they had not participated in the other experiments. The procedure was the same as in Experiments 1a and 2.

Materials and design. The story and storybook were similar to Experiments 1a and 2. The materials differed from all of our other studies in that confidence was held constant: At Time 1, each witness was equally confident about three facts. Both claimed with confidence that they knew the weather (both said it was sunny), the color of the baseball (one said it was white, the other said blue), and who hit the ball through the window (one said Kenny, the other said Tyler). As in our previous storybook experiments, participants then indicated which witness they believed (Time 1).

At Time 2, information about witness accuracy was revealed. Both witnesses turned out to be right about the weather; however, one witness was right about the color of the ball, whereas the other was wrong. Given this new information, participants were asked again which witness they believed (Time 2). As a manipulation check, participants were also asked whether each witness had always been right and answered yes or no. The order in which the witnesses testified and which witness was more confident was counterbalanced across participants.

Results and Discussion

Preliminary analyses showed no significant effects involving age or sex of the participants, order, or actor, so we collapsed across these variables. Sixteen participants out of 24 (67%) passed the manipulation check questions indicating they recognized the difference between the accuracy levels of each witness.⁴

The children were sensitive to the witnesses' accuracy. At Time 1, the witnesses were equally confident and equally accurate, so we expected no significant preference for either witness. As Figure 3 shows, of the participants who passed the manipulation checks, six participants (37.5%) chose the witness who later ended up being right about the color of the ball, and nine participants (56%) chose the other witness (one participant abstained); this preference was not different from chance, $\chi^2(1, N = 15) = 0.60$, p = .44.

At Time 2, after the error was revealed, 14 participants (87.5%) believed the witness who had been accurate about all of the details of the incident, and two (12.5%) believed the witness who was incorrect about one of the details, $\chi^2(1, N=16)=9.0, p=.003$. Whereas before errors were revealed there was no significant preference for either witness, participants afterward clearly preferred the accurate witness. Thus, children's preference for which witness to believe changed once they had information about the witnesses' accuracy.

The results of this experiment are striking because they speak to how remarkably sensitive young children are to the accuracy of their informants. In almost all previous studies investigating credibility, children have had multiple opportunities to observe whether an informant was accurate or inaccurate before trusting one of those informants. In Experiment 3, we found that 5- and 6-year-old children showed a strong preference for an informant who made no mistakes over one who made a single error about a relatively trivial detail (the color of a ball). These results also demonstrate that children's failure to use calibration information in Experiment 1a was not because they did not notice or care about the particular mistake the informant made.

Experiment 4: Cognitive Load and Adults' Use of Calibration

The results of Experiments 1–3 demonstrate that although children can maintain and use information about informants' confidence and accuracy separately when deciding whom to believe, only adults use information about informants' calibration. Making use of calibration information requires recognizing confidence level, recognizing accuracy level, and then binding the two together (e.g., Fact 1 = high confidence + right; Fact 2 = low confidence + wrong). Binding the relevant pieces of information to determine calibration may simply be too cognitively taxing for 5- and 6-year-old children (see Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Lorsbach, & Reimer, 2005), especially when multiple factors (e.g., overall confidence and calibration level) vary across informants.

An alternative possibility to deficits in children's ability to bind confidence and accuracy together is that children do not use calibration information because of deficits in retention of independent information about confidence and accuracy. Although Experiments 2 and 3 suggest that children can understand and use confidence and accuracy information separately, we can test this hypothesis in another way. Previous research has postulated that dividing attention in young adults can mimic attention and reason-

⁴ We analyzed the results regarding which witness participants believed, with and without those who failed this manipulation check. The pattern of results and significance was the same in both cases.

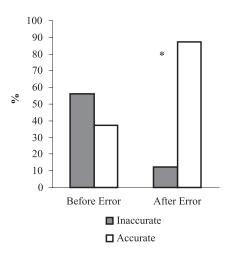


Figure 3. Percentage of participants (who passed the manipulation check questions) who chose to believe each witness before and after one witness was shown to have made an error about a detail and the other was shown to have been accurate in Experiment 3. p < 0.05.

ing deficits in other groups (Cowan et al., 2006; Sirois & Shultz, 2006). In Experiment 4, we investigated the possibility that requiring adult participants to complete a secondary task while evaluating witness credibility might impair their ability to make use of calibration information. We tested adults under cognitive load or no cognitive load while they watched video segments of a wellcalibrated and a poorly calibrated witness giving depositions about who was at fault in a car accident. We used a video about a car accident trial rather than the storybook or object naming paradigms from Experiments 1–3 because video lends itself more easily to a cognitive load manipulation and is more appropriate for an adult population. If cognitive load impairs adults' ability to make use of calibration information when judging witnesses' credibility when they are still able to keep track of confidence and accuracy, then perhaps children's failure to use calibration is due to insufficient cognitive resources to bind information together rather than an inability to maintain the independent features that comprise it.

Method

Participants. One hundred fourteen undergraduates at a selective public university (67 women, 47 men) participated in a 15-min session and received psychology course credit.

Design. This experiment was a 2 (witness confidence: confident, cautious) \times 2 (witness accuracy: Time 1 before error, Time 2 after error) \times 2 (cognitive load: yes, no) mixed factorial design, with confidence and accuracy within subjects and cognitive load between subjects.

Materials. We created a video version of the car accident stories that we have used in previous research and described in the introduction (e.g., Tenney et al., 2007). The structure of the story is analogous to the broken window story used in Experiment 1a. We paid two male actors in their late 20s to act as the eyewitnesses. They gave their depositions in a law school classroom that was designed to look like a courtroom.

Procedure. Participants were randomly assigned to the cognitive load or no cognitive load condition. Each participant sat at

a computer in a room alone with an experimenter. In the cognitive load condition, participants practiced counting backward from 1,000 by twos out loud until they felt comfortable with the task. Then, all participants read on a computer screen that they would act as jurors to determine who was at fault in an automobile accident.

At this point, participants in the cognitive load condition began counting backward aloud and watched the deposition of the first witness while they counted. In the no cognitive load condition, participants watched the deposition without counting. One witness (the confident witness) was highly confident about everything: He was positive that he went to the post office and took his dog to the veterinarian that day, and he was positive that the red car hit the gray car. The other witness (the cautious witness) was highly confident that he went to a meeting at work about remodeling, was unsure about whether he had lunch with a neighbor that day, but was certain that the gray car hit the red car.

After each witness testified, participants in the cognitive load condition were instructed to stop counting, and all participants rated each witness's credibility on a scale from 1 (*not credible*) to 6 (*credible*). After both witnesses had testified, participants were asked to choose which witness's account of the car accident they believed. The evaluations up to this point constituted Time 1 judgments.

Next, participants in the cognitive load condition began counting again. All participants saw a still picture of each witness and listened to an audio recording, which explained that both witnesses (a) had been right about one activity they had claimed occurred on the day of the accident but (b) had been proven wrong about the second activity they had claimed to have done that day—records revealed that the second activities had actually been done on different days. After hearing this new information, participants in the cognitive load condition stopped counting, and all participants again rated each witness's credibility and chose whose account of the car accident they believed. These represented the Time 2 judgments.

As manipulation checks, at the end of the study, participants also answered questions about whether the witnesses had made errors, how distracted they had felt during the experiment from 1 (not distracted) to 6 (very distracted), and what they believed was each witness's overall confidence level from 1 (not confident) to 6 (very confident).

We counterbalanced the following variables across participants: the order in which the confident and cautious witnesses testified, the actors who played the confident and cautious witnesses, and the activities and descriptions of the car accident.

Results and Discussion

Manipulation checks.

Distraction. Participants in the cognitive load condition said an average of 108 numbers during the experiment with a standard deviation of 35.5. As expected, participants in the cognitive load condition reported being more distracted (M = 5.3, SD = 0.84) than participants in the no cognitive load condition (M = 2.2, SD = 1.18), t(112) = 16.32, p < .001, d = 3.11.

Accuracy. Seven out of 55 participants in the cognitive load condition (12.7%) reported that the confident witness did not make an error. Because they failed this critical manipulation check, these

seven participants were excluded from further analyses. None of the participants in the no-load condition made this error.

Confidence. Overall, the confident witness was rated as more confident (M = 5.3, SD = 0.99) than the cautious witness (M = 3.5, SD = 1.17), F(1, 106) = 130.79, p < .001, d = 1.76. This pattern was consistent across cognitive load conditions (i.e., the interaction between confidence and cognitive load was only marginally significant), F(1, 105) = 3.58, p = .061, d = 0.60.

Actor and order. Unexpectedly, when Actor A was confident, participants were more influenced by calibration than when Actor B was confident ($d=1.40~\rm vs.~0.80$). In addition, participants were more influenced by calibration when the confident witness testified first than when he testified second ($d=1.22~\rm vs.~0.87$). To account for these random effects of which actor played the confident witness and which order the witnesses testified, we included these variables as covariates in subsequent analyses of witness credibility (although their inclusion did not end up affecting significance). There were no significant effects of participant sex, so this variable was not considered further.

Believability. Cognitive load affected participants' decisions about which witness's account of the car accident they believed. At Time 1, before errors were revealed, there was an overall effect of confidence: 76 participants (71%) chose to believe the confident witness, and only 31 (29%) chose to believe the cautious one, $\chi^2(1,$

N=107) = 18.93, p < .001. As expected, this pattern did not differ across no cognitive load (45/59 [76%] for confident) and cognitive load (31/48 [65%] for confident) conditions, $\chi^2(1, N=107)=1.76$, p=.18.

At Time 2, after errors were revealed, 40 participants (68%) in the no-load condition chose the cautious, well-calibrated witness, and 27 participants (56%) in the cognitive load condition did so. Although these percentages do not significantly differ, it is more instructive to consider how participants' responses changed from Time 1 to Time 2. As the left panels of Figure 4 show, in the no-load condition 26 participants (44%) switched from the confident to the cautious witness, which was significantly greater than would be expected by chance (i.e., 14.8 participants [25%]). No participants switched in the opposite direction. Nineteen participants (32%) chose the confident witness twice and 14 (24%) chose the cautious witness twice; neither of these two choice patterns was significantly different from what would be expected by chance. A 1 × 4 chi-square goodness-of-fit test comparing all possible combinations of Time 1 - Time 2 choices was significant, $\chi^2(3, N = 59) = 24.59$, p < .001. Thus, there was a strong shift in preference from the confident witness at Time 1 to the cautious (well-calibrated) one at Time 2 in the no-load condition.

In contrast, in the cognitive load condition, 15 (31%) switched from the confident to the cautious witness, five (10%) switched in

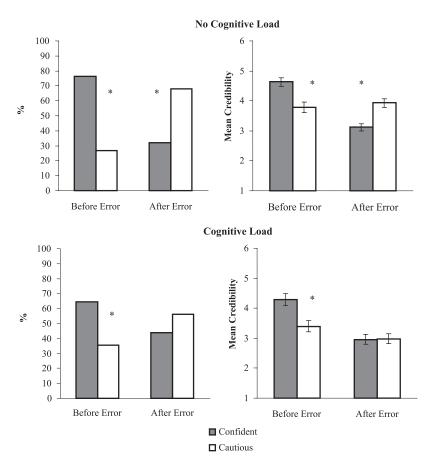


Figure 4. Percentage of participants who chose to believe the confident and cautious witnesses (left panel) and mean credibility rating (1-6) of the confident and cautious witnesses (right panel) before and after errors were revealed in Experiment 4. Error bars represent standard error of the mean. * p < .05.

the opposite direction, 16 (33%) chose the confident witness twice, and 12 (25%) chose the cautious witness twice. However, a 1 × 4 chi-square goodness-of-fit test revealed that the distribution of participants' choices across Time 1 and Time 2 was not significantly different from what would be expected by chance (i.e., 12 participants [25%]), $\chi^2(3, N=48)=6.17$, p=.104. There was no significant shift in preference from the confident to cautious (well-calibrated) witnesses.

Credibility. As the right panels of Figure 4 show, credibility ratings at Time 1 and Time 2 corroborated the believability judgments. We first conducted repeated-measures analysis of covariance on credibility separately for cognitive load and no cognitive load conditions, with confidence and accuracy as within-subject variables; actor and order were included as covariates. In the no cognitive load condition, as expected, there was a strong, significant interaction between confidence and accuracy, F(1, 56) = 25.35, p < .001, d = 1.19. This simple interaction can be interpreted to mean that the effect of making an error on credibility depended on the witness's confidence in that error (i.e., calibration). In the cognitive load condition, there was also a significant interaction, F(1, 45) = 7.79, p = .008, d = 0.75; however, the effect size of the interaction was smaller.

Second, to determine whether the interaction between confidence and accuracy was stronger in the no cognitive load condition than in the cognitive load condition, we conducted a repeated-measures analysis of covariance with cognitive load as a between-subjects factor. As predicted, the three-way interaction among confidence, accuracy, and cognitive load was statistically significant, F(1, 103) = 4.44, p = .038, implying that the moderating effect of confidence on the relationship between accuracy and credibility was larger in the no cognitive load condition.

Taken together, these results suggest that participants kept track of confidence and accuracy but relied on calibration information more when they were not under cognitive load. Thus, consistent with previous ideas that feature binding is particularly cognitively taxing (Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996; Reinitz, Morrissey, & Demb, 1994), it may be possible for children and distracted adults to retain the individual features necessary to recognize calibration information (i.e., confidence and accuracy) without connecting or binding them to use when judging credibility.

General Discussion

We hypothesized that because of disparities in cognitive abilities, children and adults might differ in the information they use to make inferences about the credibility of informants. Experiment 1 showed that adults, but not children, were sensitive to information about informants' calibration—that is, adult participants were influenced by whether informants were good judges of their own knowledge and exuded the appropriate amount of confidence. Children, on the other hand, tended to believe whichever informant was the most confident, even when the informants' confidence was overstated.

Artifactual Explanations

One potential hypothesis about the discrepancy between children and adults' use of calibration information is that the materials in these particular experiments made it too difficult for children to notice or comprehend the components that made up calibration: confidence and accuracy. However, we ruled out that explanation. First, in Experiments 1a and 1b, children showed a clear preference for the confident informant—they were not at chance, where one would expect them to be had they been overwhelmed by the task. Second, in Experiments 2 and 3, using the same storybook paradigm as Experiment 1a, children chose to believe the more confident informant and the more accurate informant when confidence and accuracy were manipulated independently. Third, Experiment 1b used a different set of materials from Experiment 1a and found the same result. Thus, children can comprehend and use the components that comprise calibration in these types of studies.

Another simple explanation to consider is that children are sensitive to calibration information, but in our particular paradigms, they continued to choose the confident informant after she was shown to be poorly calibrated because they were reluctant to change their first answer. The results of Experiment 3, however, suggest that this explanation is unlikely. In Experiment 3, children who learned that one informant was more accurate than the other had no difficulty switching their responses to the more accurate informant (see also Scofield & Behrend, 2008, who found that 4-year-olds would cease relying on an informant who was shown to be inaccurate). Thus, the artifactual explanations of children being unable to comprehend the relevant pieces of information or being reluctant to change their answers do not explain the differences between children and adults' discrepant choice of informants.

Capacity-Based Explanations

One plausible explanation for our findings is that children do not make use of calibration information because it requires them to integrate confidence and accuracy information, and they have difficulty doing so. Some evidence that children tend to rely on independent features comes from lifespan literature on feature binding in memory. Feature binding (also called cohesion or associative learning) is when aspects of a stimulus event are remembered together, like remembering someone's name and face or recalling a fact and how that fact was learned (e.g., Chalfonte & Johnson, 1996; Metcalfe, Mencl, & Cottrell, 1994; Reinitz & Hannigan, 2001). People can remember each individual feature separately and, in addition, remember the features as a combined, cohesive unit to represent the stimulus event in memory; however, binding features together takes more cognitive resources than remembering single items. Research has shown that feature binding requires more attention than single feature encoding (Kroll et al., 1996; Reinitz, Morrissey, & Demb, 1994), and children and older adults have more difficulty with feature binding than young adults (Cowan et al., 2006; Lorsbach & Reimer, 2005; Sluzenski, Newcombe, & Kovacs, 2006). Young adults under cognitive load have also shown decreased binding ability (Cowan et al., 2006). Thus, the ability to bind information together is a crucial part of the developmental process that seems to improve during childhood and decline during older adult life, as attention, working memory capacity, long term memory, and other cognitive resources wax and wane. In the current experiments, to use information about informants' calibration, participants had to combine informants' confidence and accuracy together (e.g., Fact 1 = high confidence + right; Fact 2 = low confidence + wrong). In Experiment 4, young adults under cognitive load showed a similar pattern to children: using confidence but not calibration. Thus, the development difference might be due to binding problems.

On the other hand, it is possible that children can bind information to infer calibration and did recognize that one informant in Experiments 1a and 1b was better calibrated, but they picked the confident informant anyway. One reason for that could be that they felt pragmatic pressure to select the confident one. Research with adults has shown that people will side with a confident advisor to make social interactions run more smoothly (Zarnoth & Sniezek, 1997) or because it is the social norm to do so (Curley, Yates, & Abrams, 1986). We think this explanation is unlikely, however, as children were not actually interacting with the informants; they were merely asked to make judgments about individuals in a storybook or on a video. Another explanation is that children might bind confidence and accuracy information but interpret the result differently from adults. Perhaps children believe that one instance of misplaced confidence doesn't matter and give confidence more weight as a signal for how often an informant will be correct in the future.

Self-Calibration

Some research suggests that children as young as 5 years old can use mental state terms like guess and know when they are uncertain and confident about a proposition, respectively (Cherney, 2003). But a good deal of other research suggests that children are not particularly well calibrated. For example, 11- and 12-year-old children were asked to assess how confident they were in their answers to 44 questions regarding a short film clip depicting a kidnapping event (Allwood, Knutsson, & Granhag, 2006). The children exhibited overconfidence in their memory assertions, consistently displaying high levels of confidence whether they provided accurate or inaccurate answers and overestimating the number of questions they answered correctly. Moreover, children showed a markedly poorer link between their ratings of confidence and the accuracy of their assertions compared with adults who viewed the same film clip and answered the same questions (Allwood, Granhag, & Johansson, 2003; Allwood et al., 2006).

Similarly, researchers examined the link between confidence and accuracy in suspect identifications of 10- to 14-year-old children and adults after they viewed a simulated crime (Keast, Brewer, & Wells, 2007). Although there was a significant relation between confidence and accuracy for adults, children were as likely to express confidence when choosing the wrong suspect as they were when choosing the right one. If children are unable to monitor their own behavior in such a way as to reflect an understanding of the limitations of their own knowledge, then they may be unable to process complex information about the link between confidence and accuracy in others' behavior.

Implications

Being well calibrated is useful for allotting study time where one needs it most, for assessing the veracity of one's memories, and for being honest with oneself about one's own strengths and weaknesses. To the extent that informants are well calibrated, people are better able to evaluate input and advice. Adults seem to recognize the value in well-calibrated informants. To adults, a statement made with confidence invites an expectation of accuracy, and a violation of this expectation will cast doubt on other statements by the same informant. But to children, a high confidence error does not carry the same repercussions for informants' credibility. Perhaps Mark Twain gave sage advice for interacting with children when he proclaimed, "Whatever you say, say it with conviction" ("Quotes," n.d.). But at precisely what age and under what circumstances calibration begins to trump confidence is yet unknown. There is much still to be learned about how calibration and judgments of credibility change across the lifespan.

References

- Allwood, C. M., Granhag, P. A., & Johansson, M. (2003). Increased realism in eyewitness confidence judgments: The effect of dyadic collaboration. Applied Cognitive Psychology, 17, 545–561. doi:10.1002/ acn 888
- Allwood, C. M., Knutsson, J., & Granhag, P. A. (2006). Eyewitness under influence: How feedback affect the realism in confidence. *Psychology*, *Crime*, & *Law*, 12, 25–38. doi:10.1080/10683160512331316316
- Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-yearolds spontaneously use others' past performance to guide their learning. *Cognition*, 107, 1018–1034. doi:10.1016/j.cognition.2007.12.008
- Borckardt, J. J., Sprohge, E., & Nash, M. (2003). Effects of the inclusion and refutation of peripheral details on eyewitness credibility. *Journal of Applied Social Psychology*, 33, 2187–2197. doi:10.1111/j.1559-1816.2003.tb01880.x
- Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Be-havior*, 26, 353–364. doi:10.1023/A:1015380522722
- Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, 24, 403–416. doi: 10.3758/BF03200930
- Cherney, I. D. (2003). Young children's spontaneous utterances of mental terms and the accuracy of their memory behaviors: A different methodological approach. *Infant and Child Development*, 12, 89–105. doi: 10.1002/icd.267
- Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: When is feature binding difficult? *Developmental Psychology*, 42, 1089–1102. doi:10.1037/0012-1649.42.6.1089
- Curley, S. P., Yates, J. F., & Abrams, R. A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes*, 38, 230–256. doi:10.1016/0749-5978(86)90018-X
- Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior*, 12, 41–55. doi:10.1007/BF01064273
- Harris, P. L. (2007). Trust. Developmental Science, 10, 135–138. doi: 10.1111/j.1467-7687.2007.00575.x
- Harris, P. L., & Richert, R. A. (2008). William James, "the world of sense" and trust in testimony. *Mind & Language*, 23, 536–551. doi:10.1111/j.1468-0017.2008.00354.x
- Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development*, 8, 263–283. doi:10.1080/15248370701446392
- Jaswal, V. K., McKercher, D. A., & VanderBorght, M. (2008). Limitations on reliability: Regularity rules in the English plural and past tense. *Child Development*, 79, 750–760. doi:10.1111/j.1467-8624.2008.01155.x
- Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. Psychological Science, 17, 757–758. doi:10.1111/j.1467-9280.2006.01778.x

- Keast, A., Brewer, N., & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology*, 97, 286–314. doi:10.1016/j.jecp.2007.01.007
- Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, 76, 1261–1277. doi:10.1111/ j.1467-8624.2005.00849.x
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*, 694–698. doi:10.1111/j.0956-7976.2004.00742.x
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370. doi:10.1037/0096-3445.126.4.349
- Kroll, N. E. A., Knight, R. T., Metcalfe, J., Wolf, E. S., & Tulving, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language*, 35, 176–196. doi:10.1006/jmla.1996.0010
- Lorsbach, T. C., & Reimer, J. F. (2005). Feature binding in children and young adults. *Journal of Genetic Psychology*, 166, 313–327. doi: 10.3200/GNTP.166.3.313-328
- Metcalfe, J., Mencl, W. E., & Cottrell, G. W. (1994). Cognitive binding. In D. Schacter (Ed.), *Memory systems* (pp. 368–394). Cambridge, MA: MIT Press
- Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development*, 60, 167–171. doi: 10.2307/1131082
- Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, 43, 1216–1226. doi:10.1037/0012-1649.43.5.1216
- Penrod, S. D., & Cutler, B. L. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy,* and Law, 1, 817–845. doi:10.1037/1076-8971.1.4.817
- Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, 17, 39–57. doi:10.1002/bdm.460
- Quotes. (n.d.) Retrieved from http://cmgww.com/historic/twain/about/quotes3.htm
- Reinitz, M. T., & Hannigan, S. L. (2001). Effects of simultaneous stimulus presentation and attention switching on memory conjunction errors. *Journal of Memory and Language*, 44, 206–219. doi:10.1006/jmla.2000.2727
- Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 161–168. doi:10.1037/0278-7393.20.1.161

- Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, 72, 1054–1070. doi:10.1111/1467-8624.00334
- Sabbagh, M. A., Wdowiak, S. D., & Ottaway, J. M. (2003). Do word learners ignore ignorant speakers? *Journal of Child Language*, 30, 905–924. doi:10.1017/S0305000903005828
- Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development*, 23, 278–290. doi:10.1016/ j.cogdev.2008.01.003
- Sirois, S., & Shultz, T. R. (2006). Preschoolers out of adults: Discriminative learning with a cognitive load. *Quarterly Journal of Experimental Psychology*, 59, 1357–1376. doi:10.1080/17470210500246269
- Sluzenski, J., Newcombe, N., & Kovacs, S. L. (2006). Binding, relational memory, and recall of naturalistic events: A developmental perspective. *Journal of Experimental Psychology: Learning, Memory, and Cogni*tion, 32, 89–100. doi:10.1037/0278-7393.32.1.89
- Sporer, S., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence–accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118, 315–327. doi:10.1037/0033-2909.118.3.315
- Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in observers' knowledge? *Child Development*, 62, 1334– 1351. doi:10.2307/1130810
- Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science*, 18, 46–50. doi:10.1111/j.1467-9280.2007.01847.x
- Tenney, E. R., & Spellman, B. A. (in press). Complex social consequences of self-knowledge. Social Psychological and Personality Science.
- Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology*, 44, 1368–1375. doi:10.1016/j.jesp.2008.04.006
- Thomas, J. P., & McFadyen, R. G. (1995). The confidence heuristic: A game-theoretic analysis. *Journal of Economic Psychology*, 16, 97–113. doi:10.1016/0167-4870(94)00032-6
- Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, 110, 611–617. doi:10.1037/0033-2909.110.3.611
- Zarnoth, P., & Sniezek, J. A. (1997). The social influence of confidence in group decision making. *Journal of Experimental Social Psychology*, 33, 345–366. doi:10.1006/jesp.1997.1326

Appendix

Example Pictures of Witnesses in the Storybook





Pointing indicates confidence

Shrug of the shoulders indicates uncertainty

Figure A1. Signed consent was provided for the likeness to be published in this article.

Received April 2, 2010
Revision received August 17, 2010
Accepted December 7, 2010