Journal of Experimental Psychology: General

ISSN: 0096-3445

2020, Vol. 149, No. 8, 1527–1536
http://dx.doi.org/10.1037/xge0000726

Exploring Information Use in Children's Decision-Making: Base-Rate Neglect and Trust in Testimony

Samantha Gualtieri
University of Waterloo

Daphna Buchsbaum
University of Toronto

Stephanie Denison
University of Waterloo

Classic literature in judgment and decision-making shows that when testimony information conflicts with base-rates, adults typically underuse base-rate information and rely heavily on testimony (Bar-Hillel, 1980; Lyon & Slovic, 1976; Tversky & Kahneman, 1981). Although children can use base-rates (Denison, Konopczynski, Garcia, & Xu, 2006; Kushnir, Xu, & Wellman, 2010) and testimony (Koenig & Harris, 2005) separately in their inferences, whether they show a similar tendency toward weighing testimony more heavily is unknown. Four- and 5-year-old children were asked to guess the color of a dog's collar, drawn from a group of 10 dogs (e.g., 8 blue: 2 yellow). Children were also presented with testimony about the dog's collar that was from either a previously accurate or inaccurate witness. In Experiment 1 (N = 120), children were presented with only base-rate or testimony information. They relied on base-rates at above chance levels and relied on testimony at rates that approximately matched the witness's previous accuracy. In Experiment 2 (N = 160), when base-rates and testimony were presented together and conflicted, a majority of children endorsed the color consistent with the accurate witness's testimony, neglecting base-rates. However, when presented with the inaccurate witness's testimony, children were more likely to endorse the color indicated by the base-rates. Children appear to rely on the testimony of an accurate but fallible witness, revealing that a tendency to neglect base-rates in favor of testimony emerges early in development, yet they remain sensitive to the witness's accuracy when presented with multiple sources of information.

Keywords: cognitive development, heuristics and biases, trust in testimony, base-rate neglect

Supplemental materials: http://dx.doi.org/10.1037/xge0000726.supp

In our daily lives, we are frequently in situations where multiple pieces of information should factor into our judgments and decisions. However, we sometimes forgo more comprehensive computations that involve integrating information, and instead make decisions using simpler strategies that trade off accuracy for speed and computational efficiency (Gigerenzer, 1997; Gigerenzer & Gaissmaier, 2011). In a classic test of this phenomenon, adults were tasked with identifying the color of a taxi-cab involved in a traffic accident (Bar-Hillel, 1980; Lyon & Slovic, 1976; Tversky & Kahneman, 1981). Participants were told that 85% of all cabs in the city were green and the other 15% were blue. A witness identified the cab as blue, and it was noted that the witness was accurate 80% of the time when identifying colors under viewing conditions similar to those during the accident. In their subsequent estimates, most participants reported that there was an 80% chance that the cab was blue. However, this estimation grossly neglects the base-rate of cabs in the city. According to Bayes' theorem, if base-rate and testimony information are appropriately considered, there is only a 41% chance of the cab being blue.[1] That is, there

This article was published Online First December 5, 2019.

Samantha Gualtieri, Department of Psychology, University of Waterloo; Daphna Buchsbaum, Department of Psychology, University of Toronto; Stephanie Denison, Department of Psychology, University of Waterloo.

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Stephanie Denison. We thank parents and children for participating. We would also like to thank principals and teachers at WRDSB and WRCSB, and the staff at THEMUSEUM and Royal Ontario Museum for their support. The data from this manuscript appears as Chapter 4 of Samantha Gualtieri's dissertation, located here: http://hdl.handle.net/10012/14837. These data, or parts of these data, have been presented at the Society for Research in Child Development, the Canadian Society for Brain, Behaviour and Cognitive Science, and the Canadian Developmental Psychology conferences. Data and stimuli can be found here: https://osf.io/bhwjs/.

Correspondence concerning this article should be addressed to Samantha Gualtieri, Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. E-mail: sgualtieri@uwaterloo.ca

[1] Bayes' Theorem: $\Pr(B|t = B) = \frac{\Pr(t = B|B)\Pr(B)}{\Pr(t = B|B)\Pr(B) + \Pr(t = b|G)\Pr(G)}$, where t is the witness's testimony, and B and G indicate blue and green respectively. We can compute $\Pr(B|t = B)$, the probability that it is really a blue car, given that the witness said it was blue, by substituting in the accuracy and base-rate information given in the classic problem: $\Pr(B|t = B) = \frac{(.8)(.15)}{(.8)(.15) + (.2)(.85)} \approx 0.41$.

is a 41% chance that the cab is blue once you consider the relatively low base-rate of blue cabs, and the chance that the witness accidentally misidentified one of the more common green cabs as blue. Instead of integrating base-rate and testimony information, adults appear to use the witness's accuracy as a shortcut for the full computation.

The primary goal of the current paper is to examine the developmental origins of adults' tendency to rely on testimony and neglect base-rates in these classic experiments. We presented 4- and 5-year-olds with a visual, child-friendly version of the taxi-cab problem, in which they must decide whether or not to agree with a witness when her testimony conflicts with base-rate information. Researchers have examined heuristic reasoning in childhood and thus far have most often studied children's use of base-rate information in variants of the lawyer-engineer problem (Davidson, 1995; De Neys & Vanderputte, 2011; Gualtieri & Denison, 2018; Jacobs & Potenza, 1991). In the most recent of these studies, 4- to 6-year-old children were presented with age-appropriate, visual versions of the classic task, in which base-rate information conflicted with a personality description of a particular individual (i.e., case-specific information). For instance, participants would see a base-rate that contained eight nice characters and two mean characters. They were asked to identify whether a randomly selected individual from the group was nice or mean, and were given additional information about the individual's traits and prior behavior. For instance, children heard that the individual enjoyed scaring other children and hiding another child's gifts. Although this information sounds indicative of a mean individual, it is not perfectly diagnostic and thus the base-rate of nice and mean individuals should remain relevant. In these problems, 4-year-olds trend more toward base-rate use, while 5-year-olds begin to show a preference for the case-specific personality and trait information, which is further strengthened to near ceiling-levels by 6 years of age. Thus, by age 6, children readily apply the representativeness heuristic in their decision-making: they opt to rely on case-specific information that closely matches their representation of a social group's characteristics when making an inference, which leads them to neglect relevant base-rate information (Gualtieri & Denison, 2018).

A similar developmental difference has been observed in American children's proclivity toward the fundamental attribution error. This error is indicated by a bias toward person-specific explanations of others' behavior that focus on an individual's traits and overlook the role of situational factors. By the age of 6, children endorse person-specific explanations of others' behavior (e.g., the girl did not go down the slide because she is scared) over situational explanations (e.g., the girl did not go down the slide because it was broken), similar to adults in Western societies. However, 4-year-olds are not as biased toward these person-specific explanations and instead stick more closely to the observed behavioral covariations (Seiver, Gopnik, & Goodman, 2013). Together, these experiments suggest that heuristic reasoning, which can sometimes result in ignoring or underusing relevant statistical information, strengthens during early childhood.

At first glance, it might seem surprising that younger children would stick more closely to statistical data in their decisions, particularly when older children and adults use heuristic shortcuts in lieu of these data. However, from as early as infancy, children are quite adept at using statistical data in their reasoning (Aslin, Saffran, & Newport, 1998; Denison et al., 2006; Denison, Bonawitz, Gopnik, & Griffiths, 2013; Girotto, Fontanari, Gonzalez, Vallortigara, & Blaye, 2016; Kirkham, Slemmer, & Johnson, 2002; Téglás, Girotto, Gonzalez, & Bonatti, 2007; Xu & Garcia, 2008). That is, infants expect the majority item to be sampled from a population of items, and can use this information to inform their decisions in a search task (Denison & Xu, 2010; Denison & Xu, 2014; see Rakoczy et al., 2014, and Tecwyn, Denison, Messer, & Buchsbaum, 2017, for evidence of this ability in nonhuman primates). In contrast, it may take greater verbal comprehension and fluency to become familiar with the sociocultural information that is necessary for using a representativeness heuristic or for making person-centered inferences as in cases of the fundamental attribution error. The combination of these factors might result in later use of sociocultural information, as opposed to statistical information, in judgments and decision-making, particularly when the information conflicts.

Although 4-year-olds are still developing their understanding of how stable traits might impact behavior (Boseovski & Lee, 2006; Boseovski, Chiu, & Marcovitch, 2013; Gonzalez, Zosuls, & Ruble, 2010; Liu, Gelman, & Wellman, 2007; Martin & Ruble, 2004; Trautner et al., 2005), they are quite adept at using information from social testimony in their inferences (see Koenig, Tiberius, & Hamlin, 2019, for a recent review). That is, social transmission of facts and norms is one of the most important sources of knowledge for very young children (Harris, Koenig, Corriveau, & Jaswal, 2018). By the preschool years, children can judge whether a particular speaker is a good source of knowledge by considering factors like their past accuracy, confidence, and expertise (Koenig & Harris, 2005; Koenig & Sabbagh, 2013; Mills, 2013; Pasquini, Corriveau, Koenig, & Harris, 2007; Poulin-Dubois & Brosseau-Liard, 2016; Sobel & Kushnir, 2013). Further, a recent review of this literature suggests that children are particularly sensitive to situational constraints that influence the value of using a person's testimony (Koenig et al., 2019). Children use factors such as a person's perceptual access and the overall plausibility of the errors they make when deciding whether, and under what circumstances, to rely on them in the future.

Given children's early emerging ability to skeptically evaluate testimony (Harris et al., 2018; Koenig et al., 2019; Mills, 2013), to make simple statistical inferences with base-rates (Xu & Garcia, 2008), and to integrate testimony with causal frequency information (Bridgers, Buchsbaum, Seiver, Griffiths, & Gopnik, 2016), we examined children's inferences when given base-rate and testimony information that conflicted. We tested these inferential abilities with 4- and 5-year-old children for two main reasons. First, 4-year-olds, but not 3-year-olds, have the ability to make rational inferences with probabilistic testimony data (Koenig & Harris, 2005; Pasquini et al., 2007). Thus, this is the youngest age group that possesses the requisite abilities to reason about testimony information in a taxi-cab-type problem. It is critical that the witnesses in the current problems are *probabilistically* accurate as in the classic adult experiments. If the witness is perfectly accurate (i.e., 100% correct) or inaccurate (i.e., 0% correct), then there is no reason to integrate testimony and base-rate information, because children should always trust a perfectly accurate witness, or mistrust a perfectly inaccurate witness.[2] Second, including 5-year-olds allows us to examine whether either integrating testimony and base-rates, or relying on testimony over base-rates (and related base-rate neglect), changes with age over this period or remains mostly stable.

Connecting the heuristics and biases and selective trust literatures has important implications for dual-process theories of cognition. These theories posit that decision-making can rely on two types of processing: Type I processing, which is relatively quick and computationally efficient, and Type II processing, which is slower and computationally expensive. Traditionally, researchers have argued that Type II processing is desirable because judgments typically consider all available information (Stanovich, West, & Toplak, 2011). Under this view, if provided with both a base-rate and witness testimony in a taxi-cab problem, reasoners should integrate these sources, rather than using a Type I shortcut of solely relying on testimony. Thus, it is possible that given 4-year-olds' strong abilities to make inferences with both statistical information and testimony information, they will integrate these sources of information. However, using heuristics can be valuable as well, because they are often effective, with the trade-off of introducing some systematic errors. Therefore, young children might use a heuristic or shortcut and rely exclusively on the testimony, given the computational efficiency. In any case, applying a heuristic in inappropriate circumstances would be entirely ineffective. It would not be useful to trust a person's testimony in cases where they have proven unreliable in the past, particularly if other high-quality information is available. Thus, examining the circumstances in which children might rely on testimony information over base-rates is pivotal to understanding children's reasoning in these situations.

In two experiments, we explored how 4- and 5-year-old children use testimony and base-rate information in tandem. The current paper builds on the emerging literature on children's judgment and decision-making, which has thus far examined children's use of the representativeness heuristic when base-rate information is pitted against case-specific information. Given that young children are more adept at using testimony information than trait information in their inferences, we extend these investigations to examine how children make judgments and decisions that involve witness testimony.

Experiment 1

Experiment 1 explored children's use of base-rate and testimony information separately in three between-subjects conditions (the base-rate condition, the accurate testimony condition, and the inaccurate testimony condition) to assess baseline use of this information for later comparisons to Experiment 2. The base-rate condition presented children with a group of 10 dogs, eight wearing one color collar and two wearing another color. We were interested in children's use of this numerical information when guessing the collar color of an unknown dog that was randomly sampled from the group. Based on previous work using similar types of paradigms, we predict that most children will choose the majority color (Denison et al., 2013; Gualtieri & Denison, 2018).

There were also two accuracy conditions. Children in both accuracy conditions were introduced to a girl who liked to watch dogs in the park and identified the colors of six dogs' collars as they caught a ball. Her accuracy at identifying colors differed across conditions: in the accurate condition, she was correct 5/6 times on the previous day, while in the inaccurate condition, she was correct 3/6 times on the previous day. Following the accuracy sequence, children were introduced to a dog whose collar color was unknown, and the girl provided testimony regarding which color she thought she saw. Children were then asked to make an inference about the color of the collar.

We developed these novel accuracy conditions to facilitate later comparisons to Experiment 2 when base-rate and testimony information is presented together. We predicted that most children should endorse the witness's testimony in the accurate condition, though it is unclear how they might use her testimony in the inaccurate condition. In some studies, children have opted to rely on information provided by an inaccurate informant at above chance levels when it is the only available information, and thus there was no conflicting information from another informant to rely on (Bridgers et al., 2016; Vanderbilt, Heyman, & Liu, 2014). In other work, children have relied on the testimony of informants at levels that approximately reflect the witness's previous accuracy (Reifen Tagar, Federico, Lyons, Ludeke, & Koenig, 2014). In the inaccurate condition of Experiment 1, the witness is correct 50% of the time, which would result in approximately 50% of children endorsing her testimony if children respond at levels consistent with her previous accuracy.

Method

Participants. This research, submitted under the name "Learning and conceptual development in infants and children" (protocol number: 30215), received ethics clearance through the University of Waterloo's Research Ethics Committee. Informed consent was obtained from guardians for all child participants. In all experiments, children were individually tested at schools in Southwestern Ontario or at a local museum. Demographic information was not formally collected, but the region is predominantly middle-class, and approximately 81% of residents in this region are Caucasian, with Chinese and South Asians as the most visible minorities (Statistics Canada, 2017).

Prior to data collection, we established the criteria that we would stop testing children after we had obtained a full sample of 40 in each condition (see Table 1 for age and gender breakdown of participants in each condition). One hundred twenty children were included in the final analyses, with 20 4-year-olds and 20 5-year-olds in each of three conditions. Six additional children were tested and excluded due to parental report of low English language exposure ($n = 3$) or noncompliance ($n = 3$).

Materials and procedure. For sample materials for both experiments, please see: https://osf.io/bhwjs/. In three between-subjects conditions, children were told a story about a girl at a dog park via a PowerPoint presentation that was narrated live by an experimenter (see Figure 1 for an overview of the procedure).

In the base-rate condition, participants saw that there were 10 dogs at the park wearing blue or yellow collars. Of the 10 dogs, eight wore one color (e.g., blue), and two wore the other color (e.g., yellow). The experimenter counted the dogs and pointed out

[2] For a 100% accurate witness, $\Pr(B|t = B) = \frac{(1)(.15)}{(1)(.15) + (0)(.85)} = 1$, so the correct behavior is always to disregard the base-rate.

Table 1
Age and Gender Breakdown per Condition in Experiment 1

Condition	Mean age	Female
Base-rate	60.58 months	18
Accurate condition	60.98 months	21
Inaccurate condition	60.75 months	20

that more dogs were wearing one of the colors. Children were then asked to indicate which color there was more of, and, depending on the child's response, the experimenter agreed or disagreed with their choice and stated that there were lots of dogs wearing blue and less wearing yellow. Children were then introduced to a dog at the park that day who was running away with a blanket covering its collar. Thus, the dog's group membership was unknown. Children were asked to recall which color there was more of, and, depending on the child's response, the experimenter agreed or disagreed with their choice. The experimenter asked the child, "What color is this one wearing?" The color introduced first, the color of majority collar, and the placement of the dogs in the base-rate array were counterbalanced.

In the accuracy conditions, participants were told that a girl at the park liked to identify what color each dog was wearing while the dog chased a ball. During the history phase, participants saw what color the girl thought she saw, followed by the actual color of each dog, for six dogs. The witness was accurate 5/6 times in the accurate condition, and 3/6 times in the inaccurate condition. Children were asked if the witness was good or not good at identifying colors. Depending on the child's response, the experimenter agreed or disagreed with their choice: the experimenter stated the girl was good because she got five right and only one wrong (accurate), or stated that she was not very good because she got three right and three wrong, and was guessing (inaccurate). Children were then introduced to a dog at the park who was running away with a blanket covering its collar. Children were told what color the girl thought the dog was wearing (i.e., "She saw it, so she says it's wearing yellow"). After this, participants were asked to recall what color the girl thought the dog was wearing and if she was good or not very good at identifying the colors before. Children were corrected if they misremembered this information. The experimenter then asked the child, "What color is this one wearing?" The color introduced first, the order of collar colors during the accuracy portion, the order of the witness's correct responses during the history phase, and the color of the witness's testimony were counterbalanced.

Results

Data for Experiments 1 and 2 can be found here: https://osf.io/bhwjs/. Children were given a score of 1 if they chose the group that was indicated by the information they were given. That is, in the base-rate condition, children were given a score of 1 if they chose the majority group, and children in the accuracy conditions were given a score of 1 if they chose the color indicated by the witness.

We examined the base-rate condition separately from the accuracy conditions, given that children in this condition were responding to the question based on different information (see Table 2 and Figure 2 for the means per condition). To explore any potential effect of age on responses, we conducted a logistic regression with children's age group (4-year-olds, 5-year-olds) in the model, which indicated no significant effects of age on performance[3], *Wald's* $\chi^2(df = 1) = .143$, $p = .71$. Overall, children chose the majority color at a rate higher than chance ($M = .78$, $SD = .42$, $p = .001$, exact binomial test) [4].

We then examined performance in the two accuracy conditions together to explore any potential effect of age. A logistic regression with accuracy condition (accurate, inaccurate) and children's age (4-year-olds, 5-year-olds) in the model revealed no significant effects of condition, *Wald's* $\chi^2(df = 1) = 1.394$, $p = .24$, or age, *Wald's* $\chi^2(df = 1) = .510$, $p = .47$. Despite the lack of condition effect, we explored children's responses in each condition to establish the extent to which they relied on the testimony when it was the only available information. In the accurate condition, children chose the group indicated by the witness at a rate higher than chance ($M = .73$, $SD = .45$, $p = .006$, exact binomial test), while performance in the inaccurate condition was not statistically different from chance ($M = .60$, $SD = .50$, $p = .26$, exact binomial test).

Discussion

To establish children's baseline behavior in our paradigm, Experiment 1 presented 4- and 5-year-old children with base-rate and testimony information separately. We observed no differences in performance as a function of children's age. Children in the base-rate condition relied on the 8:2 base-rate information and selected the majority group in their inferences at rates higher than chance. In the testimony conditions, children's responses did not significantly differ based on the witness's accuracy. It appears that children in each of the testimony conditions used the witness's testimony at rates roughly reflecting her prior accuracy (similarly to Reifen Tagar et al., 2014). Children presented with an accurate witness used her testimony at rates above chance, and children presented with an inaccurate witness used her testimony at rates close to 50%, which corresponds to both chance and her previous accuracy level.

Experiment 2

The results of Experiment 1 provide context for interpreting children's responses when they are presented with base-rate and testimony information together in the same problem. We manipulated the witness's accuracy at identifying colors (accurate: correct 5/6 times; inaccurate: correct 3/6 times) and whether this aligned or conflicted with the base-rate of dogs (no conflict: her

[3] For all regression analyses across both experiments, we found similar effects (no changes in significance cut-offs) when age was treated continuously.

[4] We also explored children's performance when they misremembered the information before the test question. Importantly, all children were corrected before moving on. In the base-rate condition, 6/40 participants misremembered the base-rate. In the accurate condition, 4/40 kids misremembered the witness's accuracy and 5/40 misremembered her testimony. In the inaccurate condition, 3/40 kids misremembered the witness's accuracy and 2/40 misremembered her testimony. Given that these numbers are so small, we did not perform any statistics, but it appears that children's data were very similar to the rest of the group when this information was misremembered but then corrected.

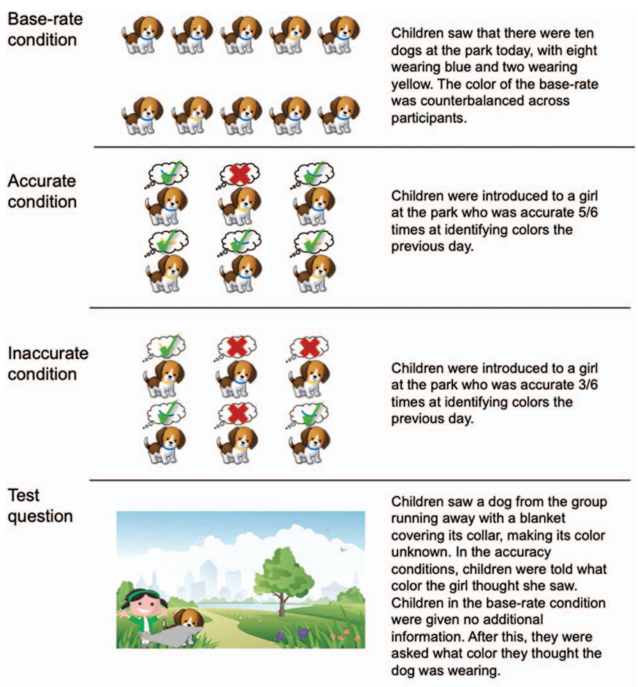

Figure 1. Overview of procedure in Experiment 1. See the online article for the color version of this figure.

testimony aligns with the majority; conflict: her testimony conflicts with the majority, as she states it is the minority color) in a 2 × 2 between-subjects design. This design results in four between-subjects conditions: the accurate conflict condition, the inaccurate conflict condition, the accurate no conflict condition, and the inaccurate no conflict condition.

The accurate conflict condition corresponds with the classic taxi-cab problem, which is why it is critical for examining children's information use. The witness, who is approximately 83% accurate, thinks that the collar of the missing dog is, for example, yellow, although 80% of the dogs are wearing blue. According to

Table 2
Children's Use of Base-Rate and Testimony Information in Each Condition

Condition	Base-rate choices		Testimony choices	
	n	%	*n*	%
Experiment 1				
Base-rate condition	31	78%		
Accurate condition			29	73%
Inaccurate condition			24	60%
Experiment 2				
Accurate no conflict	38	95%	38	95%
Accurate conflict	6	15%	34	85%
Inaccurate no conflict	36	90%	36	90%
Inaccurate conflict	26	65%	14	35%

Note. n = 40 per condition. In Experiment 1, children were given either only testimony or only base-rate information. In Experiment 2, the base-rate and testimony information cued opposite responses in the conflict conditions, but cued the same response in the no conflict conditions.

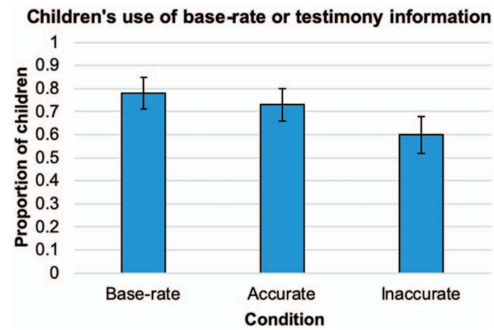

Figure 2. Proportion of children choosing the higher base-rate option in the base-rate condition, and the testimony option in the accurate and inaccurate conditions. See the online article for the color version of this figure.

Bayes' theorem, if children integrate the base-rate information with the witness's accuracy, they should say that the dog is wearing blue 45% of the time, as a group. If they instead mostly rely on the witness's testimony, then they should say the dog is wearing yellow approximately 83% of the time.

In the inaccurate conflict condition, the witness, who has been correct just 50% of the time, believes that the collar is yellow, and 80% of the dogs are wearing blue. This condition examines whether children elect to use the reliable information (i.e., the base-rate information) rather than the testimony information when a witness has proven to be unreliable. If children entirely neglect base-rates in favor of testimony, even when the witness has a history of inaccuracy, then it is possible they will use her testimony at a rate similar to Experiment 1 (i.e., approximately 60% of the time).

The two no conflict conditions serve as reference points for children's performance in this more complicated task. In the inaccurate no conflict condition, the witness is only 50% accurate and states that the collar is blue when 80% of the dogs are also wearing blue. This condition is included to rule out the possibility that children may reflexively provide the *opposite* response to an inaccurate witness's testimony, regardless of base-rates, when the problem becomes more complex and potentially harder to follow. Employing a shortcut to simply give the opposite response to the inaccurate witness would be irrational in this situation because the base-rate information points in the same direction. The accurate no conflict condition should be entirely uncomplicated. The witness, who is correct 83% of the time, thinks that the collar is blue and 80% of the dogs are also wearing blue. In sum, children in both no conflict conditions should choose the color endorsed by the witness and the base-rate information. These conditions also allow us to assess whether having two converging pieces of information have an additive effect on children's decisions.

Method

Participants. We again tested 40 children in each condition. 160 children were included in the final analyses, with 20 4-year-olds and 20 5-year-olds in each of the four conditions (see Table 3 for age and gender breakdown). Five additional children were tested and excluded because of interruption in the testing environ-

Table 3
Age and Gender Breakdown per Condition

Condition	Mean age	Female
Accurate no conflict	60.90 months	24
Accurate conflict	60.37 months	19
Inaccurate no conflict	60.65 months	21
Inaccurate conflict	60.80 months	24

ment (contractors entered the room during testing; $n = 1$) or noncompliance ($n = 4$).

Materials and procedure. Participants were told that a girl at the park liked to identify what color collar each dog was wearing while they chased a ball (see Figure 3 for an overview of the procedure and online supplemental materials for sample stimuli). During the history phase, participants were told about the witness's accuracy when identifying the colors of six dogs on the previous day, using the same 5/6 or 3/6 accuracy rates as in the accuracy conditions for Experiment 1. Participants then saw a group of 10 new dogs and were told that these dogs were at the park on the current day. The experimenter counted the dogs and established the majority (8:2) as in the base-rate condition of Experiment 1. Because children were presented with two pieces of information in Experiment 2, we included a recap slide where the experimenter reminded participants what color there was more of at the park on the current day, and how accurate the witness was at identifying colors the previous day. Information was always recapped in this order, mimicking the structure of the typical adult taxi-cab problem. This recap reduced the memory demands of the task and replaced the questions that the experimenter previously asked the children (and corrected if they provided incorrect responses). Because children's performance did not differ based on whether they misremembered this information or remembered correctly in Experiment 1, these questions were replaced with this recap slide to shorten the procedure while still ensuring that all children were reminded of the correct information. Children were then introduced to a dog at the park that day who was running away with a blanket covering its collar, making its group membership unknown. Children were told what color the girl thought the dog was wearing (i.e., "She saw it, so she says it's wearing yellow"). The experimenter then asked the child, "What color is this one wearing?". The color introduced first, the order of collar colors during the accuracy portion, the order of the witness's correct responses during the history phase, the color of the majority collar, the placement of the dogs in the base-rate array, and the color of the witness's testimony were counterbalanced.

Results

Children received a score of 1 if they selected the group indicated by the base-rate (in no conflict cases, this cues the same response as when coded by testimony). See Figure 4 for a graph of the means per condition.

To explore children's responses across conditions and any effects of age, we conducted a logistic regression with conflict condition (conflict, no conflict), accuracy condition (accurate, inaccurate), children's age (4-year-olds, 5-year-olds), and the interaction between conflict condition and accuracy condition included in the model. This revealed a significant main effect of conflict condition, *Wald's* $\chi^2(df = 1) = 35.273$, $p < .001$, and an interaction between conflict and accuracy condition, *Wald's* $\chi^2(df = 1) = 8.662$, $p = .003$, no main effect of accuracy condition, *Wald's* $\chi^2(df = 1) = 2.325$, $p = .12$, and no main effect of age, *Wald's* $\chi^2(df = 1) = 0$, $p = 1$. The interaction was driven by children's performance in the conflict condition; children's use of base-rate information on conflict problems significantly differed based on the witness's accuracy ($p < .001$, Fisher's exact test).

To further examine children's use of base-rate and testimony information, we compared children's performance in Experiment 2 to the baseline conditions in Experiment 1 (see Table 2 for a comparison of performance). We first explored children's performance in the no conflict conditions. Testimony and base-rate information cued the same group in the no conflict conditions, and thus higher scores reflect a tendency to respond based on both types of information. These responses were then compared to the baseline base-rate and testimony performance in Experiment 1. In the accurate no conflict condition, children's responses ($M = .95$, $SD = .22$) differed significantly from their base-rate use in Experiment 1 ($M = .78$, $SD = .42$; $p = .048$, Fisher's exact test), and their use of testimony in the accurate condition in Experiment 1 ($M = .73$, $SD = .45$; $p = .013$, Fisher's exact test). This suggests that when the information converges and all information is reliable and relevant, there is an additive effect on children's judgments.

In the inaccurate no conflict condition, children's responses ($M = .90$, $SD = .30$) did not differ significantly from their base-rate use in Experiment 1 ($M = .78$, $SD = .42$; $p = .23$, Fisher's exact test). However, children's responses differed significantly from their use of testimony in the inaccurate condition of Experiment 1 ($M = .60$, $SD = .50$; $p = .004$, Fisher's exact test). In this case, having the reliable base-rate information coupled with the unreliable testimony led children to make stronger inferences than with unreliable testimony alone. Overall, the results from the no conflict conditions confirm that participants could follow the narrative in both accuracy conditions, and that they do not automatically disagree with an inaccurate witness.

We then examined children's performance in the conflict conditions, in which testimony and base-rate information cued different colors. We first examined performance in the accurate conflict condition, which maps onto the classic taxi-cab problem. First, we examined if children's responses were in line with an integration strategy that normatively weighs both base-rate and accurate testimony information. If children were using this strategy, approximately 45% of participants should choose the group cued by the base-rate. We found that their performance significantly differed from this value ($M = .15$, $SD = .36$; $p < .001$, exact binomial test). We then examined whether children might be relying only or primarily on testimony by comparing their performance to the accurate testimony condition of Experiment 1, where they received only testimony information. Children relied on the testimony information in Experiment 2 ($M = .85$, $SD = .36$, coding reversed for comparison, i.e., in Experiment 2, the 15% base-rate use is equivalent to 85% testimony use) at similar rates to Experiment 1 ($M = .73$, $SD = .45$; $p = .27$, Fisher's exact test), suggesting that they were focusing on this information in Experiment 2. Altogether, these analyses are most consistent with the interpretation that, when presented with a conflict between an accurate witness

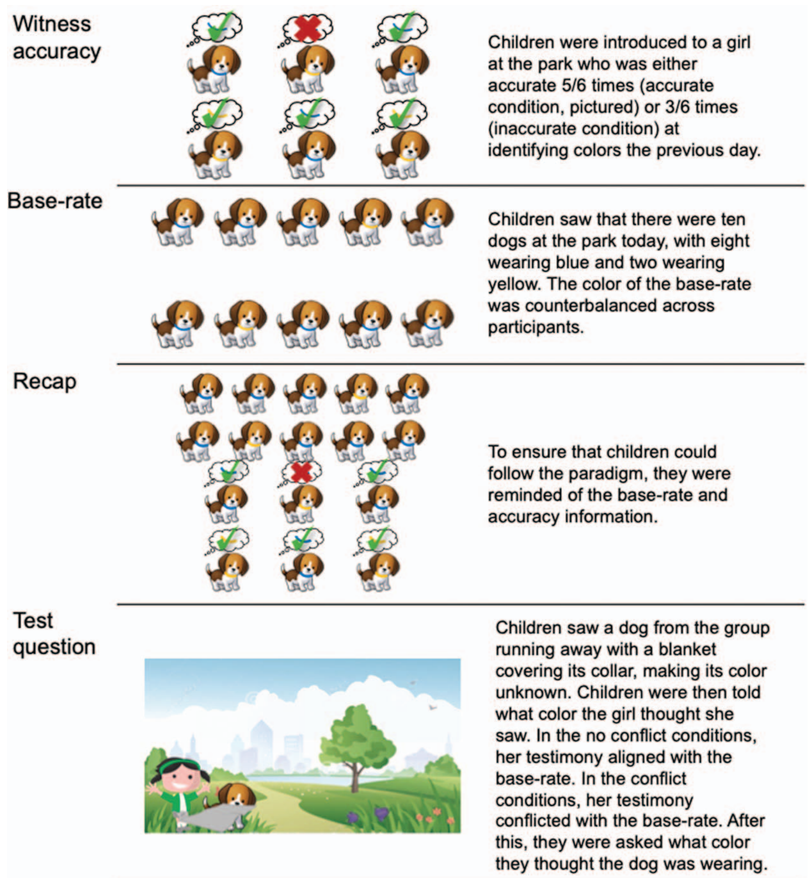

Figure 3. Overview of procedure in Experiment 2. See the online article for the color version of this figure.

and base-rate information, children did not integrate base-rates and testimony but instead neglected base-rates.

Finally, we explored children's performance in the inaccurate conflict condition. In this condition, the inaccurate testimony conflicted with more reliable base-rate information, so reliance on testimony in this context would be ineffective. Children's use of testimony information (M = .35, SD = .48, reverse coded) differed significantly from their use of testimony in Experiment 1, as they relied on the witness significantly more in their inferences in Experiment 1 (M = .60, SD = .49; p = .043, Fisher's exact test).

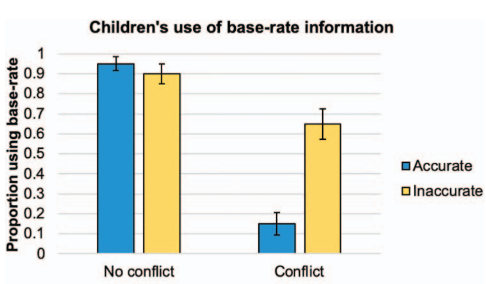

Figure 4. Proportion of children choosing the higher base-rate option in each condition. See the online article for the color version of this figure.

Thus, children were selective in their use of testimony when more reliable base-rate information was available.

Discussion

In Experiment 2, we presented children with problems in which the base-rate and testimony information either aligned or conflicted. When both pieces of information aligned, children performed at near-ceiling levels, selecting the color indicated by both the base-rate and the witness. Children chose the color that was cued by both pieces of information more often than when either piece was presented alone in Experiment 1, suggesting that there was an additive effect when the information was reliable. However, children relied heavily on the accurate witness's testimony when it conflicted with the base-rate, opting to use the testimony to make inferences about the collar color. Notably, this preference to rely on testimony was not extended to the inaccurate witness. Whereas relying on an accurate witness who claims to have had perceptual access to an event is reasonable, relying on the testimony of a previously inaccurate witness would be irrational when other information is available. When the inaccurate witness's testimony conflicted with the base-rates, children were pulled more toward the base-rate information and did not reflexively rely on the testimony information.

General Discussion

The current experiments explored how children reconcile information from witness testimony with base-rates and found that children had a selective preference for testimony information when these two sources are in conflict. At baseline, children relied on either base-rate or accurate testimony information when they were presented separately in Experiment 1. When given testimony from the inaccurate witness, children reasonably considered her low prior accuracy of 50% and did not rely on her testimony at rates higher than chance. In Experiment 2, children were presented with base-rates and testimony information together. When both pieces of information supported the same inference, children performed at near ceiling levels and selected the group that was suggested by both the base-rates and testimony. The pivotal conflict conditions presented children with conflicting base-rate and testimony information. When the witness was accurate, children were more likely to use her testimony in their inferences than the base-rates, and group level responses indicated no signs of integrating the evidence. This is very similar to adult behavior in the classic problem; adult judgments differ from the value that would be predicted if base-rates and testimony were integrated, and do not differ from the value that would be predicted if only witness accuracy was considered. This suggests that a preference to rely on testimony from an accurate source emerges early in development. However, children's preference for testimony information over base-rates was selective. Compared to the accurate conflict condition, children in the inaccurate conflict condition were more likely to select the color that was cued by the base-rates. Further, children endorsed the inaccurate witness's testimony in the inaccurate conflict condition at a rate significantly lower than in the inaccurate baseline condition of Experiment 1. This suggests that when testimony from an inaccurate witness conflicts with the reliable information from the base-rate, children appropriately place more weight on the base-rate information. Taken together, these findings suggest that children can use a testimony shortcut at very young ages, but they do so selectively.

Much previous work has established that young children reliably use base-rates (Denison et al., 2006; Kushnir et al., 2010; Ma & Xu, 2011) and accurate testimony information (Harris et al., 2018; Koenig & Harris, 2005; Pasquini et al., 2007) in their inferences. The current findings suggest that a tendency to favor testimony over base-rates is present by 4 years of age, with young children preferentially relying on the information provided by an accurate, but imperfect, witness rather than conflicting base-rates. An interesting question is whether this finding should be interpreted as evidence of a "bias" for testimony as in the classic adult heuristics and biases literature, which prescribes integration of these sources of information as the mathematically correct solution. Although thinking of this as a bias is a reasonable interpretation of children's performance, a second, equally reasonable interpretation is that children's behavior is quite rational, despite deviating from mathematical normativity. That is, children were informed of the answer to a question by a witness who had visual access to the event, and rather than spending a great deal of cognitive energy integrating base-rates and accuracy, they elected to trust her. This aligns with recent interpretations of the selective trust literature, which would also predict that children should rely on the witness because the situational constraints remained constant from her previous performance and she is stated to have had visual access to the event (Koenig et al., 2019). This behavior is also consistent with theories of bounded rationality and resource-rational inference. In contrast to dual process theories of cognition, these positions argue that human decision-makers often make predictions that are "boundedly" optimal within the constraints of their cognitive systems, trading off precision for more efficient decision-making strategies (Gigerenzer, 1997; Gigerenzer & Gaissmaier, 2011; Lieder & Griffiths, 2019).

Regardless of whether the behavior of children in our studies should be interpreted as rational or not, or how rationality should be defined, the findings of the current experiments are helpful in understanding the development of heuristic use and base-rate neglect. Notably, young children did not rely on the testimony of the inaccurate witness to the same extent as the accurate witness when her testimony conflicted with the base-rates. Children in the inaccurate testimony condition gave more base-rate consistent responses than those in the accurate testimony condition. Recent findings, in which children were presented with a single, inaccurate informant, have also found that children's use of inaccurate testimony is contingent on the presence of conflicting information, which may facilitate their ability to weigh and contrast the information they are given (Bridgers et al., 2016; Vanderbilt et al., 2014). Children opted to rely on information provided by an inaccurate informant when it was the only available piece of information. However, children relied on a neutral informant, with no prior history of accuracy, who provided information that conflicted with the inaccurate informant (Vanderbilt et al., 2014). In the current experiments, children agreed with the inaccurate witness at a rate that matched quite closely to her accuracy level of 50% in Experiment 1. Similar to previous findings, when the inaccurate witness was paired with more reliable base-rate information, children trusted the inaccurate witness less. Together with other recent findings on children's ability to integrate testimony and causal frequency information (Bridgers et al., 2016), and with the additive effect of these factors in the accurate no conflict condition of Experiment 2, these findings suggest that young children can effectively weigh testimony information with other pieces of information.

Limitations and Future Directions

In order to be accessible to young children, we used a forced-choice response method in our design. In the classic adult paradigm, participants are asked to rate the likelihood that the taxi-cab is blue, as the witness said (Bar-Hillel, 1980; Lyon & Slovic, 1976; Tversky & Kahneman, 1981). We used a binary choice paradigm to ensure that 4- and 5-year-old participants were able to provide a response, because children this age cannot estimate likelihoods using percent values or provide relevant explanations for their thought processes. A binary response is also desirable from an ecological validity perspective because, regardless of certainty, people often ultimately have to make categorical decisions. Nonetheless, future studies could employ a rating scale to obtain more sensitive and graded judgments, providing additional insight into children's degree of belief in a particular choice. Previous work with young children has indicated that an individual child's responses over repeated trials tend to reflect the group distribution as a whole, suggesting that aggregating responses across a group of children in a forced-choice paradigm reliably represents an individual child's beliefs (Denison et al., 2013).

In addition, our stimuli were presented to children in a visual format, rather than as a written-out story with numerical values, and thus may have been more likely to engender frequency-based representations of the information. Findings from the adult judgment and decision-making literature have shown that participants are more likely to make use of base-rate information in cases where the stimuli are presented as frequencies instead of percentages (Gigerenzer & Hoffrage, 1995; Hoffrage, Krauss, Martignon, & Gigerenzer, 2015; Zhu & Gigerenzer, 2006). Since presenting word problems that contain percent values is not feasible when testing 4- and 5-year-olds, we are cautious about comparing these findings to the classic adult literature. For instance, it is possible that this same format of stimuli presentation would encourage greater base-rate use in adults than was seen in classic experiments. We are currently pursuing questions of whether visual stimuli presentations, such as the ones used here, will result in more base-rate use, or better integration in adult samples.

Conclusion

The current study is the first to explore 4- and 5-year-old children's use of base-rates in the presence of conflicting testimony information. Though young children elected to rely on the testimony of an accurate witness when it conflicted with base-rates, they were selective in their use of inaccurate testimony. Because young children are quite sophisticated in their use of testimony and base-rate information early in development, the current findings have important implications for the development of heuristic thinking in children.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321–324. http://dx.doi.org/10.1111/1467-9280.00063

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44,* 211–233. http://dx.doi.org/10.1016/0001-6918(80)90046-3

Boseovski, J. J., Chiu, K., & Marcovitch, S. (2013). Integration of behavioral frequency and intention information in young children's trait attributions. *Social Development, 22,* 38–57. http://dx.doi.org/10.1111/sode.12008

Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology, 42,* 500–513. http://dx.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,* 9–18. http://dx.doi.org/10.1037/a0039830

Davidson, D. (1995). The representativeness heuristic and the conjunction fallacy effect in children's decision making. *Merrill-Palmer Quarterly, 41,* 328–346.

De Neys, W., & Vanderputte, K. (2011). When less is not always more: Stereotype knowledge and reasoning development. *Developmental Psychology, 47,* 432–441. http://dx.doi.org/10.1037/a0021313

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The Sampling Hypothesis. *Cognition, 126,* 285–300. http://dx.doi.org/10.1016/j.cognition.2012.10.010

Denison, S., Konopczynski, K., Garcia, V., & Xu, F. (2006). Probabilistic reasoning in preschoolers: Random sampling and base rate. In R. Sun and N. Miyake (Eds.) *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1216–1221). New York, NY: CRC Press.

Denison, S., & Xu, F. (2010). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science, 13,* 798–803. http://dx.doi.org/10.1111/j.1467-7687.2009.00943.x

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition, 130,* 335–347. http://dx.doi.org/10.1016/j.cognition.2013.12.001

Gigerenzer, G. (1997). Bounded rationality: Models of fast and frugal inference. *Swiss Journal of Economics and Statistics, 133,* 201–218.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62,* 451–482. http://dx.doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684–704. http://dx.doi.org/10.1037/0033-295X.102.4.684

Girotto, V., Fontanari, L., Gonzalez, M., Vallortigara, G., & Blaye, A. (2016). Young children do not succeed in choice tasks that imply evaluating chances. *Cognition, 152,* 32–39. http://dx.doi.org/10.1016/j.cognition.2016.03.010

Gonzalez, C. M., Zosuls, K. M., & Ruble, D. N. (2010). Traits as dimensions or categories? Developmental change in the understanding of trait terms. *Developmental Psychology, 46,* 1078–1088. http://dx.doi.org/10.1037/a0020207

Gualtieri, S., & Denison, S. (2018). The development of the representativeness heuristic in young children. *Journal of Experimental Child Psychology, 174,* 60–76. http://dx.doi.org/10.1016/j.jecp.2018.05.006

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. http://dx.doi.org/10.1146/annurev-psych-122216-011710

Hoffrage, U., Krauss, S., Martignon, L., & Gigerenzer, G. (2015). Natural frequencies improve Bayesian reasoning in simple and complex inference tasks. *Frontiers in Psychology, 6,* 1473. http://dx.doi.org/10.3389/fpsyg.2015.01473

Jacobs, J. E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development, 62,* 166–178. http://dx.doi.org/10.2307/1130712

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), B35–B42. http://dx.doi.org/10.1016/S0010-0277(02)00004-5

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

Koenig, M. A., & Sabbagh, M. A. (2013). Selective social learning: New perspectives on learning from others. *Developmental Psychology, 49,* 399–403. http://dx.doi.org/10.1037/a0031619

Koenig, M. A., Tiberius, V., & Hamlin, J. K. (2019). Children's judgments of epistemic and moral agents: From situations to intentions. *Perspectives on Psychological Science, 14,* 344–360. http://dx.doi.org/10.1177/1745691618805452

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science, 21,* 1134–1140. http://dx.doi.org/10.1177/0956797610376652

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*. Advance online publication. http://dx.doi.org/10.1017/S0140525X1900061X

Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development, 78,* 1543–1558. http://dx.doi.org/10.1111/j.1467-8624.2007.01082.x

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica, 40,* 287–298. http://dx.doi.org/10.1016/0001-6918(76)90032-9

Ma, L., & Xu, F. (2011). Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition, 120,* 403–411. http://dx.doi.org/10.1016/j.cognition.2011.02.003

Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science, 13,* 67–70. http://dx.doi.org/10.1111/j.0963-7214.2004.00276.x

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology, 49,* 404–418. http://dx.doi.org/10.1037/a0029500

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

Poulin-Dubois, D., & Brosseau-Liard, P. (2016). The developmental origins of selective social learning. *Current Directions in Psychological Science, 25,* 60–64. http://dx.doi.org/10.1177/0963721415613962

Rakoczy, H., Clüver, A., Saucke, L., Stoffregen, N., Gräbener, A., Migura, J., & Call, J. (2014). Apes are intuitive statisticians. *Cognition, 131,* 60–68. http://dx.doi.org/10.1016/j.cognition.2013.12.011

Reifen Tagar, M., Federico, C. M., Lyons, K. E., Ludeke, S., & Koenig, M. A. (2014). Heralding the authoritarian? Orientation toward authority in early childhood. *Psychological Science, 25,* 883–892. http://dx.doi.org/10.1177/0956797613516470

Seiver, E., Gopnik, A., & Goodman, N. D. (2013). Did she jump because she was the big sister or because the trampoline was safe? Causal inference and the development of social attribution. *Child Development, 84,* 443–454. http://dx.doi.org/10.1111/j.1467-8624.2012.01865.x

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

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31,* 103–118. http://dx.doi.org/10.1016/j.dr.2011.07.003

Statistics Canada. (2017). *Census Profile, 2016 Census: Waterloo, Regional municipality [Census division], Ontario and Ontario [Province]* (No: 98-316-X2016001). Retrieved from https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E

Tecwyn, E. C., Denison, S., Messer, E. J., & Buchsbaum, D. (2017). Intuitive probabilistic inference in capuchin monkeys. *Animal Cognition, 20,* 243–256. http://dx.doi.org/10.1007/s10071-016-1043-9

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 19156–19159. http://dx.doi.org/10.1073/pnas.0700271104

Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development, 14,* 365–381. http://dx.doi.org/10.1002/icd.399

Tversky, A., & Kahneman, D. (1981). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). New York, NY: Cambridge University Press.

Vanderbilt, K. E., Heyman, G. D., & Liu, D. (2014). In the absence of conflicting testimony young children trust inaccurate informants. *Developmental Science, 17,* 443–451. http://dx.doi.org/10.1111/desc.12134

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 5012–5015. http://dx.doi.org/10.1073/pnas.0704450105

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition, 98,* 287–308. http://dx.doi.org/10.1016/j.cognition.2004.12.003

Received May 17, 2019
Revision received September 13, 2019
Accepted October 28, 2019 ■