A description–experience gap in statistical intuitions: Of smart babies, risk-savvy chimps, intuitive statisticians, and stupid grown-ups

Christin Schulze [*], Ralph Hertwig

Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

ABSTRACT

Comparison of different lines of research on statistical intuitions and probabilistic reasoning reveals several puzzling contradictions. Whereas babies seem to be intuitive statisticians, surprisingly capable of statistical learning and inference, adults' statistical inferences have been found to be inconsistent with the rules of probability theory and statistics. Whereas researchers in the 1960s concluded that people's probability updating is "conservatively" proportional to normative predictions, probability updating research in the 1970s suggested that people are incapable of following Bayes's rule. And whereas animals appear to be strikingly risk savvy, humans often seem "irrational" when dealing with probabilistic information. Drawing on research on the description–experience gap in risky choice, we integrate and systematize these findings from disparate fields of inquiry that have, to date, operated largely in parallel. Our synthesis shows that a key factor in understanding inconsistencies in statistical intuitions research is whether probabilistic inferences are based on symbolic, abstract descriptions or on the direct experience of statistical information. We delineate this view from other conceptual accounts, consider potential mechanisms by which attributes of first-hand experience can facilitate appropriate statistical inference, and identify conditions under which they improve or impair probabilistic reasoning. To capture the full scope of human statistical intuition, we conclude, research on probabilistic reasoning across the lifespan, across species, and across research traditions must bear in mind that experience and symbolic description of the world may engage systematically distinct cognitive processes.

1. Introduction

Why are grown-ups often so stupid about probabilities when even babies and chimps can be so smart?

— Alison Gopnik

As developmental psychologist Gopnik (2014) noted in an article written for *The Wall Street Journal*, juxtaposing distinct lines of research on statistical intuitions reveals several puzzling inconsistencies. On the one hand, recent studies on infant cognition have shown that babies possess a remarkable ability to absorb, process, and apply everyday statistical information. In their first year of life, infants are capable of drawing accurate inferences from samples to populations (e.g., Denison & Xu, 2014; Xu & Garcia, 2008), integrating physical information about objects when making statistical inferences (e.g., Denison & Xu, 2010a; Téglás, Girotto, Gonzalez, & Bonatti, 2007), and taking into account information about the intentions of sampling agents (e.g., Xu & Denison,

2009). Likewise, recent research on animal cognition suggests that several nonhuman primates and some parrot species share human infants' ability to make adequate statistical inferences (e.g., Bastos & Taylor, 2020; De Petrillo & Rosati, 2019; Eckert, Rakoczy, Call, Herrmann, & Hanus, 2018; Rakoczy et al., 2014; Tecwyn, Denison, Messer, & Buchsbaum, 2017). On the other hand, research conducted in the heuristics-and-biases tradition has found that adults' statistical inferences are often inconsistent with the rules of probability theory and statistics (e.g., Tversky & Kahneman, 1974). For nearly five decades, this research tradition has profoundly shaped scholars' conception of adults' statistical intuitions in psychology, economics, and beyond, concluding that "intuitive judgments of all relevant marginal, conjunctive, and conditional probabilities are not likely to be coherent, that is, to satisfy the constraints of probability theory" (Tversky & Kahneman, 1983, p. 313). Yet just a few years before the emergence of heuristics-and-biases research, a seminal review of studies on adult statistical inference—tellingly titled "Man as an intuitive statistician"—had come to the opposite conclusion: that "probability theory and statistics can be used

as the basis for psychological models that integrate and account for human performance in a wide range of inferential tasks" (Peterson & Beach, 1967, p. 29). In short, several decades of research from different lines of inquiry have reached surprisingly inconsistent conclusions, despite addressing a common question: How good are humans' (and other animals') statistical intuitions?

In this article, we aim to systematize and synthesize this body of research, bringing together research areas that have, to date, operated largely in parallel. We argue that a key factor in understanding incongruities in research on statistical intuitions (though, of course, not the only one, see 4. Other Possible Contributors to Inconsistencies in Research on Statistical Intuitions) is whether the experimental paradigms used to measure these intuitions offer authentic experience of the relevant probabilistic information or descriptive summary representations thereof. By experience, we mean situations that involve interaction with the environment to reach an understanding of its statistical structure—for instance, by sampling information sequentially, making repeated judgments with or without feedback, or actively operating or observing the experimental microworld. In description, by contrast, a full or partial symbolic summary representation of the situation is provided (see Hertwig, Hogarth, & Lejarraga, 2018). To appreciate this difference, consider how experimenters typically communicate with adults and with babies or nonhuman animals. One of the greatest cultural achievements of humankind is the capacity to communicate by means of written symbols, a powerful form of self-expression that makes the spoken word permanent and enables humans to draw on the accumulated knowledge of others (Schmandt-Besserat, 1996). Infants do not yet possess this ability to process and produce symbolic representations of the world. Therefore, whereas the experimental protocols used with adults typically involve symbolic (often text-based) description of information, the protocols used with infants and, similarly, with animals tend to involve first-hand experience of information.

Description and experience, even if structurally isomorphic, have been shown to elicit potentially different processes and behaviors: Both children and adults show systematically different choices and choice processes in description-based monetary gambles than they do when they learn about outcomes and their relative frequencies from experience (e.g., Barron & Erev, 2003; Erev, Ert, Plonsky, Cohen, & Cohen, 2017; Hertwig, Barron, Weber, & Erev, 2004; Plonsky, Teodorescu, & Erev, 2015; Rakow & Rahim, 2010; Weber, Shafir, & Blais, 2004; for reviews, see Hertwig & Erev, 2009; Rakow & Newell, 2010; for a meta-analysis, see Wulff, Mergenthaler-Canseco, & Hertwig, 2018). Recently, researchers have also found evidence for such *description–experience gaps* in other domains, such as intertemporal choice (Dai, Pachur, Pleskac, & Hertwig, 2019), social interaction in strategic games (Martin, Gonzalez, Juvina, & Lebiere, 2014), decisions under ambiguity (Dutt, Arló-Costa, Helzner, & Gonzalez, 2014; Güney & Newell, 2015), online consumer choice (Wulff, Hills, & Hertwig, 2015), financial risk taking (Lejarraga, Woike, & Hertwig, 2016), medical judgments and decisions (Armstrong & Spaniol, 2017; Lejarraga, Pachur, Frey, & Hertwig, 2016; Wegier & Shaffer, 2017), confidence estimates (Camilleri & Newell, 2019; Lejarraga & Lejarraga, 2020), adolescent risk taking (van den Bos & Hertwig, 2017), categorization (Nelson, McKenzie, Cottrell, & Sejnowski, 2010), causal reasoning (Rehder & Waldmann, 2017), and visual search (Zhang & Houpt, 2020). Moreover, recent research has contrasted perceptuo-motor decision making, a form of experience-based choice in which participants gauge relevant probabilities from their ability to perform a perceptual or motor decision task, with decision making in described lotteries (e.g., Glaser, Trommershäuser, Mamassian, & Maloney, 2012; Jarvstad, Hahn, Rushton, & Warren, 2013; Trommershäuser, Maloney, & Landy, 2008; Wu, Delgado, & Maloney, 2009).

In nearly all cases, the description–experience dimension has been studied within a single developmental period—typically adulthood. We propose that the distinction between description and experience may be one key to understanding otherwise puzzling behavioral patterns across development. Furthermore, we argue that disparities in research

findings on animal versus human cognition—and indeed in findings on probabilistic reasoning in research from the 1960s versus 1970s—can be better understood by considering the degree of experiential involvement inherent in the respective experimental protocols. In the following, we first review findings from diverse domains of probabilistic reasoning research and highlight how distinguishing between description and experience can forge bridges between these largely unrelated fields, reconciling their seemingly inconsistent findings. Second, we discuss empirical evidence supporting a significant role of the description–experience dimension in determining people's statistical intuitions. Third, we delineate our framework from other views on inconsistencies in probabilistic reasoning research. Fourth, we summarize characteristics that define experience, identify the mechanisms by which they facilitate appropriate statistical inference, and consider the conditions under which attributes of experience can improve or impair probabilistic reasoning. Finally, we outline implications for the development of broadly informed paradigms in statistical intuitions research, for theory integration, and for educational applications.

2. Description–Experience Gaps in Research on Statistical Intuitions

2.1. Statistical Intuitions From Infancy to Adulthood

2.1.1. Babies' intuitive understanding of statistics

A recent surge in studies on infants' probabilistic reasoning (for a review, see Denison & Xu, 2019) has fueled the development of novel paradigms for testing their statistical intuitions. Fig. 1 illustrates one such experimental paradigm, in which colored objects are sequentially sampled from a box, the content of the box is revealed, and infants' looking time is measured. Based on the premise that infants look at unexpected or novel events for longer than at expected or known events, several studies have found that babies form adequate expectations about the underlying population from the randomly drawn sample (Denison, Reed, & Xu, 2013; Denison & Xu, 2010a; Xu & Denison, 2009; Xu & Garcia, 2008). Likewise, infants have been shown to consider the properties of a population when inferring the likelihood that a single- or a multi-object sample will be drawn from that population (Kayhan, Gredebäck, & Lindskog, 2018; Téglás et al., 2007; Xu & Garcia, 2008). Moreover, infants use statistical information to anticipate probabilistic outcomes (Téglás & Bonatti, 2016), to make inductive generalizations (Gweon, Tenenbaum, & Schulz, 2010), to differentiate between the relative likelihoods of two simultaneous events (Kayhan et al., 2018), and to guide their exploratory actions (Sim & Xu, 2017) and preferential choices (Denison & Xu, 2010b, 2014). Experiments using this last preferential choice task have shown that infants use proportional rather than quantitative reasoning to make statistical inferences (Denison & Xu, 2014)—two possibilities that are confounded in the setup illustrated in Fig. 1. Infants also appear to integrate information about the physical properties of objects (e.g., mobility, location, density, and speed of movement) when making statistical inferences (Denison, Trikutam, & Xu, 2014; Denison & Xu, 2010a; Lawson & Rakison, 2013; Téglás et al., 2007; Téglás, Ibanez-Lillo, Costa, & Bonatti, 2015; Téglás, Vul, Girotto, Gonzalez, Tenenbaum, & Bonatti, 2011). For example, if some balls in the paradigm illustrated in Fig. 1 can be sampled without constraints whereas others are fixated and cannot be sampled, infants disregard the immovable subset when making statistical inferences (Denison et al., 2014; Denison & Xu, 2010a).

Further, infants are sensitive to the randomness of the sampling process and integrate attributes of the sampling agent (e.g., expressed preferences or visual access) when making inferences about object populations and inductive generalizations about object properties (Attisano & Denison, 2020; Gweon et al., 2010; Xu & Denison, 2009). They can infer other people's preferences from the statistical properties of their actions (Kushnir, Xu, & Wellman, 2010; Ma & Xu, 2011; Wellman, Kushnir, Xu, & Brink, 2016) and attribute the observation of
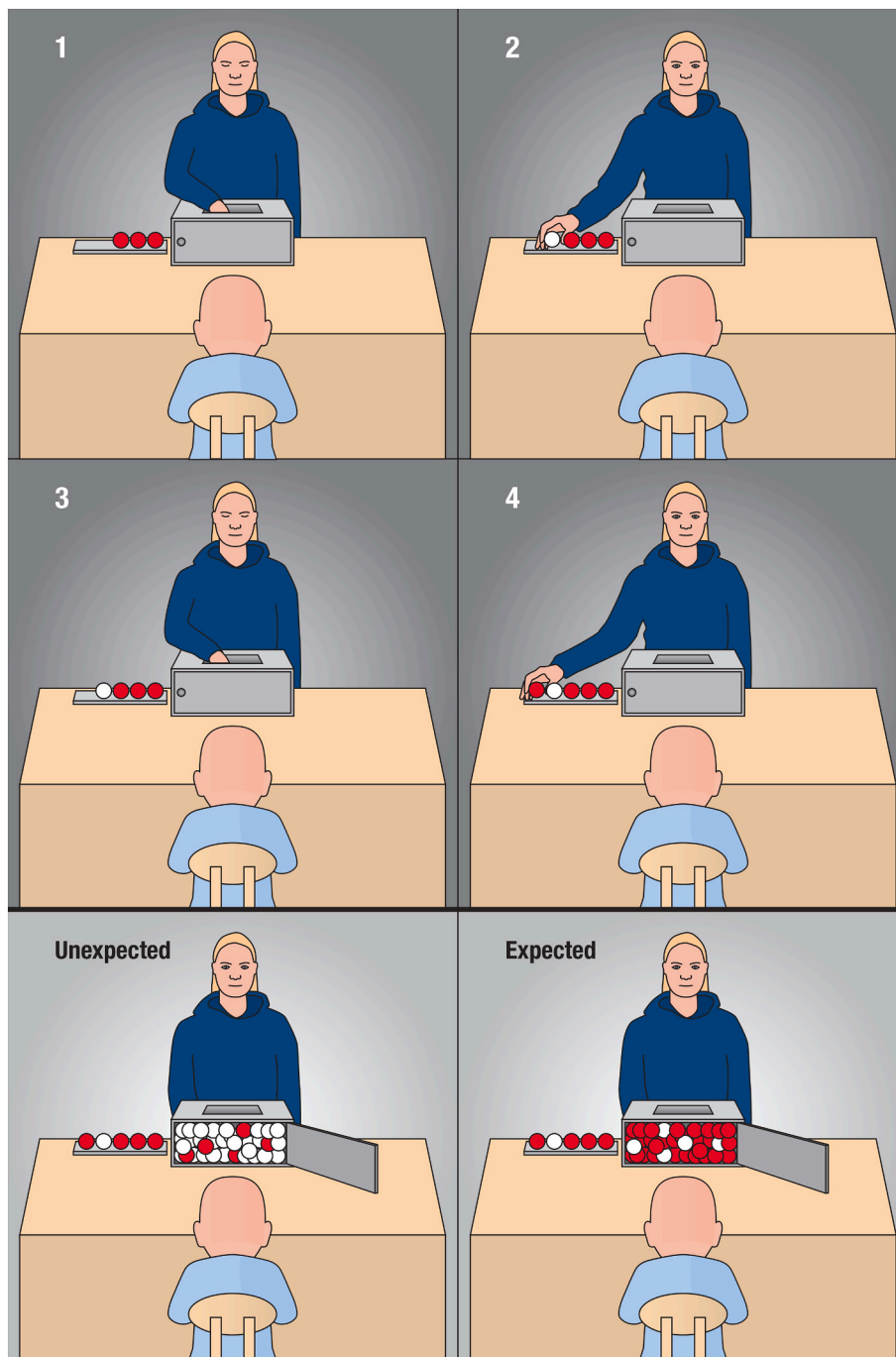
Fig. 1. Illustration of one experimental paradigm used in research on infants' statistical intuitions (e.g., Xu & Garcia, 2008). The experimenter draws several colored balls (e.g., red or white) from an opaque box without looking, the content of the box is revealed, and infants' looking time is recorded. If the sample drawn does not reflect the object distribution in the box (unexpected outcome, bottom left panel), infants tend to look at the contents of the box for longer than if the sample is consistent with the object distribution (expected outcome, bottom right panel). Adapted from Gopnik (2012).

nonrandom samples to intentional agency (Ma & Xu, 2013). Infants also use statistical information to infer the cause of a failed action: When given a dysfunctional toy, infants attribute its failure to work to either the toy or another person, depending on whether or not the toy's properties differ from those of the other person's—functional—toy (Gweon & Schulz, 2011). Table 1 summarizes this research on infants' statistical intuitions, describing the experimental protocol used in each study.

In addition, research on statistical learning in infancy has shown that babies are sensitive to statistical patterns in environmental input and use this information to learn about language (Aslin, Saffran, & Newport, 1998; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Aslin, & Newport, 1996), musical tones (Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Reeck, Niebuhr, & Wilson, 2005), and visual stimuli

(Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002). Taken together, research on statistical inference and statistical learning in infancy suggests that babies are already equipped with the ability to learn, infer, and generalize from finite environmental samples. Importantly, statistical intuitions appear to emerge irrespective of formal education and culture: Both prenumerate indigenous adults and formally educated Western controls make suitable probabilistic predictions (Fontanari, Gonzalez, Vallortigara, & Girotto, 2014).

2.1.2. Adults' error-prone statistical intuitions

Results from heuristics-and-biases research paint a different picture. Research in this tradition has identified numerous errors—often called *cognitive illusions*—that adults tend to commit when judging the probability of an uncertain event, gauging the value of an uncertain quantity,

Table 1
Summary of Experience-Based Procedures Used in Studies Testing Infants' Statistical Intuitions.

Paradigm and study	Infants' mean age (in months)	Sampling materials	Experience-based procedure and measurement	Physical instantiation
Single sample from population				
Denison and Xu (2010b)	13.5	Different ratios of pink and black lollipops in two jars (Denison & Xu, 2014; Exp. 3–4: additional green "neutral" lollipops)	1 lollipop was drawn from each population (sequentially) and infants made a single forced choice	Direct interaction
Denison and Xu (2014)	11			
Gweon et al. (2010): Exp. 2	15.5	Blue and yellow balls in one box	1 blue ball was drawn from the population and an action demonstrated (ball squeezed to make it squeak); 1 yellow ball was drawn and the infant's object manipulation coded	Direct interaction
Lawson and Rakison (2013)	8.5 or 12	Red and blue balls in two boxes with inverse ratios (e.g., 4:1 and 1:4)	1 ball was released from each population (simultaneously); VoE	Movies on screen
Téglás et al. (2007): Exp. 1–2	12	Yellow and blue geometrical objects (ratio 3:1) moving inside a transparent container with one opening	Objects moved inside a container until one exited through an opening; VoE	Movies on screen
Téglás et al. (2011)	12			
Téglás et al. (2015)	12.5			
Téglás and Bonatti (2016)	12.5	A striped ball moving inside a rectangular frame with 3 openings on one vertical side and 1 opening on the opposite side	The ball moved inside the frame until it exited at the side with 3 openings (probable outcome) or 1 opening (improbable outcome); eye-tracking and VoE	Movies on screen
Repeated sampling from population				
Attisano and Denison (2020)	6.5 or 9.5	Different ratios of red balls and yellow rubber ducks in one box	One toy was grasped repeatedly by either the experimenter or a mechanical claw in up to 14 habituation trials; at test, 3 toys of the same type were sampled by either the experimenter with visual access to the population or a mechanical claw, the population was then revealed; VoE	Puppet stage
Denison et al. (2013)	4.5 or 6	Different ratios of yellow and pink ping-pong balls in two boxes	5 balls were drawn from one of two populations in two sequential actions (3, 2 balls at a time) and both populations then revealed; VoE	Puppet stage
Denison et al. (2014)	11	Different ratios of red and green (plus yellow in Denison & Xu, 2010a; Exp. 3) ping-pong balls in (what appeared to be) two boxes. Some or all balls of one type were fixated inside the box with Velcro strips	5 balls were drawn from one of two populations in two sequential actions (3, 2 balls at a time) and the population was then revealed; VoE	Puppet stage
Denison and Xu (2010a)	11			
Gweon and Schulz (2011)*	16	Green, yellow, and red toys that played music	Activation of a toy was demonstrated repeatedly, infants failed to activate a toy of the same or a different color, and their change action (change agent or change toy) was coded	Direct interaction
Gweon et al. (2010): Exp. 1–5	15.5	Different ratios of blue and yellow balls in one box	2 or 3 blue balls were sampled sequentially (one at a time) from the population, an action was demonstrated (ball squeezed to make it squeak); 1 yellow ball was sampled and the infant's object manipulation coded	Direct interaction
Kayhan et al. (2018)	6, 12, or 18	Different ratios of colored balls (e.g., blue and orange) in one container	Two samples, one likely and one unlikely with respect to the population ratio, were simultaneously released from the container (sample size: 6 or 8 balls); eye-tracking	Movies on screen
Kushnir et al. (2010): Exp. 2	20	Different ratios of yellow rubber ducks and green rubber frogs in one box	5 toys of the same type were sampled sequentially (one at a time) and infants' choice behavior was coded ("Just what I wanted! Can you give me one?")	Direct interaction
Ma and Xu (2011)	16.5 or 26.5	Different ratios of boring (e.g., white wooden cubes) and interesting (e.g., orange slinkies with pumpkin-face print) toys in one jar	6 boring toys were sampled sequentially (one at a time) and infants' choice behavior was coded ("Can I have the one I like?")	Direct interaction
Ma and Xu (2013)	9.5	Yellow and red ping-pong balls (ratio 2:1) in one jar	9 balls were sampled sequentially (one at a time) in each sampling event by either a human hand or mechanical claw; 3 sampling events per trial; VoE	Puppet stage
Sim and Xu (2017)	13	Six balls of different colors (red, purple, blue, green, yellow, and orange) in one box (VoE) or each of two boxes (choice task)	4 balls of either the same or different colors were sampled sequentially (one at a time) with replacement; VoE or exploratory choice ("Do you want to come and play?")	Puppet stage
Wellman et al. (2016)	10	Different ratios of blue and red balls in one box	5 balls of the same color were drawn repeatedly from the population in three sequential actions (2,2,1 ball[s] at a time) in up to 8 habituation trials; at test, the experimenter selected one type of ball; VoE	Puppet stage
Xu and Denison (2009)	11	Different ratios of red and white ping-pong balls in (what appeared to be) two boxes	5 balls were drawn sequentially from one of two populations (either in two actions, Xu & Denison, 2009; or one at a time, Xu & Garcia, 2008) and the population was then revealed. In Xu and Denison (2009), whether or not sampling was random was varied (with or without visual access to the population during sampling). In Xu and Garcia (2008; Exp. 4–6), the population was revealed before the 5 balls were sampled; VoE	Puppet stage
Xu and Garcia (2008)	8			

Note. In addition to the experimental procedures being experience based, in all but two studies (Gweon et al., 2010; Kayhan et al., 2018), infants were also familiarized with (i.e., had prior exposure to) the sampling materials and sampling procedure used.
VoE: Violation-of-expectation looking paradigm.
* This study differs from the others in that it involved sampling of "actions" (i.e., repeated demonstrations) rather than repeated sampling of objects.

Table 2

Summary of Procedures Used to Demonstrate Biases in Adults' Statistical Intuitions, as Reviewed in Tversky and Kahneman (1974).

Bias and task used to demonstrate it	Example procedure and study
Insensitivity to prior probability of outcomes	
Ranking of outcomes based on likelihood versus similarity	Text-based instruction to rank nine fields of graduate specialization in order of the likelihood that a hypothetical person described in a personality sketch is a student in these fields (Kahneman & Tversky, 1973)
Judgment of the likelihood of outcomes in the presence of prior versus individuating evidence	Engineer–lawyer problem: Text-based instruction to judge the probability that a thumbnail description of a professional individual, sampled at random from a population of 30 engineers and 70 lawyers (or vice versa), belongs to one of the engineers in the population (Kahneman & Tversky, 1973)
Insensitivity to sample size	
Production of sampling distributions for different sample sizes	Text-based instruction to produce a sampling distribution of the average height of men examined at a regional induction center with seven categories (up to 160 cm, 160–165 cm, …, more than 185 cm) and for different sample sizes (N = 10, 100, or 1000; Kahneman & Tversky, 1972)
Judgment of the likelihood of sampling outcomes contingent on sample size	Maternity-ward problem: Text-based instruction to judge which of two hospitals, a smaller or a larger one, is more likely to have recorded a higher number of days on which over 60% of babies born were boys (Kahneman & Tversky, 1972)
Judgment of posterior probability for different sample sizes	Text-based instruction to judge the posterior probability that samples of marked cards in different sample sizes and sample ratios (e.g., 5 cards marked X and 1 card marked O vs. 40 cards marked X and 20 cards marked O) were drawn from a deck of cards marked either mostly X or mostly O (Kahneman & Tversky, 1972)
Misconceptions of chance	
Judgment of the likelihood of a sequence of random events	Text-based instruction to judge the relative likelihood of birth orders of boys and girls in families with six children (e.g., B G B B B B vs. G B G B B G; Kahneman & Tversky, 1972)
Judgment of the likelihood of the reproducibility of results in samples of different sizes	Text-based instruction to experienced research psychologists to judge the probability of a successful replication of a significant result in a sample smaller than the one that produced the original finding (Tversky & Kahneman, 1971)
Insensitivity to predictability	
Numerical prediction of a remote criterion versus evaluation of inputs	Text-based instruction to judge the professional standing of school teachers based on descriptive summaries of their performance in a practice lesson held 5 years earlier (Kahneman & Tversky, 1973)
The illusion of validity	
Confidence in prediction from consistent versus inconsistent evidence	Text-based instruction to report confidence in predicting grade point averages based on pairs of aptitude tests that produced either consistent scores and were described as correlated or inconsistent scores and were described as uncorrelated (Kahneman & Tversky, 1973)
Misconceptions of regression	
Anticipation and interpretation of regression toward the mean	Text-based instruction to interpret a description of a flight maneuver training situation in which verbal rewards for successes resulted in subsequent performance loss (Kahneman & Tversky, 1973)
*Biases due to the retrievability of instances**	
Judgment of the frequency of a class based on the familiarity versus occurrence rate of its instances	Sequential presentation of a recorded list of names of more or less famous men and women and verbal instruction to judge whether the list contained more men's names or women's names (Tversky & Kahneman, 1973)
Biases due to the effectiveness of a search set	
Judgment of word frequency	Text-based instruction to judge the relative likelihood with which various letters of the alphabet (e.g., K, L, N, R, V) appear in the first and third positions in words in the English language (Tversky & Kahneman, 1973)
Biases of imaginability	
Estimation of the numerosity of possible combinations	Text-based instruction to estimate the number of all possible distinct committees of various sizes (between 2 and 8 members) that can be formed from a set of 10 people (Tversky & Kahneman, 1973)
*Illusory correlation**	
Judgment of the frequency of co-occurrence of events	Sequential presentation of a recorded list consisting of highly related (knife–fork) and unrelated (head–fork) word pairs. Half the pairs were repeated three times in the recording; the other half, twice. Text-based instruction to judge the frequency with which each word pair was presented (from a written list of all pairs; e.g., Tversky & Kahneman, 1973)
Insufficient adjustment	
Estimation of quantities relative to an uninformative number	Verbal instruction to estimate the percentage of African countries in the UN relative to an arbitrary number between 0 and 100 generated by spinning a wheel of fortune in the participant's presence (Tversky & Kahneman, 1974)
Estimation of quantities based on an incomplete computation	High school students estimated, within 5 seconds, the product of a numerical expression written on the blackboard (either 8 × 7 × 6 × 5 × 4 × 3 × 2 × 1 or 1 × 2 × 3 × 4 × 5 × 6 × 7 × 8; Tversky & Kahneman, 1973)
Biases in the evaluation of conjunctive and disjunctive events	
Choice among gambles with elementary, conjunctive, and disjunctive events	Repeated choice between described pairs of gambles, one elementary (draw of one marble from a mixed urn) and one compound, either conjunctive (multiple successive draws of the same color marbles) or disjunctive (multiple successive draws in which a particular color marble is drawn at least once). At the end of the experiment, one gamble chosen by the participant was played out (Bar-Hillel, 1973)
Anchoring in the assessment of subjective probability distributions	
Elicitation of subjective probability distributions	Text-based instruction to assess the 10th or 90th percentile in the distribution of subjective beliefs about the air distance from New Delhi to Peking (group 1); or to assess the odds that the median air distance from New Delhi to Peking given by group 1 exceeds the true value (Tversky & Kahneman, 1974)

* Cognitive biases demonstrated with experience-based experimental protocols.

or deciding between risky prospects. For instance, adults commit the conjunction fallacy, do not properly account for sample sizes, overestimate the prevalence of easily recalled events, fail to make full use of base-rate information, do not anticipate regression toward the mean, are susceptible to a problem's reference frame, are overconfident, and violate various axioms of expected utility theory (Fischhoff, Slovic, & Lichtenstein, 1977; Kahneman & Tversky, 1972, 1973, 1979; Tversky & Kahneman, 1973, 1974, 1981, 1983).

What explains this inconsistency between infants' and adults' statistical inference skills? Gopnik (2014) suggested that infants' "intuitive, unconscious statistical ability may be completely separate from [adults'] conscious reasoning." This perspective echoes a dual-system view frequently invoked in heuristics-and-biases research—except that there, reasoning errors are typically attributed to the intuitive system (Kahneman, 2011; Kahneman & Frederick, 2002), which Gopnik sees as the very engine of infants' statistical competence (see also Oaksford & Hall, 2016). Several explanations are conceivable and it is likely that more than one factor is at play. One explanation that has not yet been thoroughly considered is the striking and systematic difference in the experimental protocols used in studies with adults versus infants. As summarized in Table 1, participants in infant studies typically experience the probabilistic texture of the experimental microworld at first hand. In the paradigm illustrated in Fig. 1, for instance, the child observes each sample being drawn one-by-one. Participants in adult studies, in contrast, typically receive description-based and symbolic task representations. In fact, nearly all classic tasks that demonstrate cognitive illusions in adults' statistical inferences are text based. No experiential engagement with raw data is required. For instance, adults' responses to the Linda task have been interpreted as evidence that their reasoning is at odds with the conjunction rule, "[p]erhaps the simplest and the most basic qualitative law of probability" (Tversky & Kahneman, 1983, p. 293). In this task, the conjunction rule is embedded within a description of a hypothetical woman and her possible occupations and avocations. Many other statistical inference problems and choice tasks—such as the engineer–lawyer problem, the maternity ward problem, the Asian disease problem (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1981), as well as monetary lotteries giving rise to various violations of expected utility theory (Kahneman & Tversky, 1979)—likewise use a description-based approach. In their foundational article in *Science*, Tversky and Kahneman (1974) summarized over a dozen putative cognitive biases observed in studies using a variety of tasks, as listed in Table 2. All but two of those tasks (see asterisks in Table 2) were solely description based.

2.2. Are People Intuitive Statisticians or Not Statisticians at All?

Notably, Tversky and Kahneman's (1974) article, which introduced the notion of heuristics-and-biases, was published just 7 years after Peterson and Beach (1967) reviewed several decades of experimental research (more than 160 experiments) on adult statistical inference—and concluded that the normative benchmarks of probability theory and statistics indeed provide a good initial description of people's statistical abilities: that people are *intuitive statisticians*. Like Lejarraga and Hertwig (2021), we suggest that this inconsistency across research traditions may also be attributable, at least in part, to differences in the experimental protocols used. Heuristics-and-biases research marked a turning point not only in the conceptualization and evaluation of people's ability to reason probabilistically but also in experimental protocols (see also Hertwig et al., 2018). Consider, for instance, research on Bayesian reasoning—the pinnacle of statistical inference and a key building block in classic models of rational choice—conducted in these two influential research traditions. In the 1970s, research conducted in the heuristics-and-biases tradition, primarily using text-based scenarios to present all available information (see top panel of Fig. 2), concluded that Bayes's theorem failed to describe the workings of the mind: "[i]n his evaluation of evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all" (Kahneman & Tversky, 1972, p. 450). The term "conservative Bayesian" harks back to Edwards (1968), who had used experience-based protocols that required sequential updating of probability estimates (see lower panel of Fig. 2) and concluded that people's probability updating, albeit "conservative" (beliefs are revised less strongly than prescribed by Bayes's theorem), was proportional to the normative values (Edwards, 1968; Edwards, Lindman, & Phillips, 1965).

These conflicting views and the shift in experimental protocols are surprising. In a quantitative assessment of the methodological protocols used in more than 600 empirical studies, Lejarraga and Hertwig (2021) recently demonstrated that heuristics-and-biases research established a new experimental protocol, with a strong emphasis on described scenarios, replacing the past emphasis on experiential experimental

---

Summary description

Consider two urns, denoted A and B, which are filled with a large number of colored balls. Urn A contains 70% red balls and 30% blue balls. Urn B contains 30% red balls and 70% blue balls.

One urn has been selected by chance, and 12 colored balls have been drawn at random from it, of which 8 are red and 4 are blue.

What do you think the probability is that the 12 colored balls were drawn from urn A, that is, from the urn in which most of the colored balls are red?
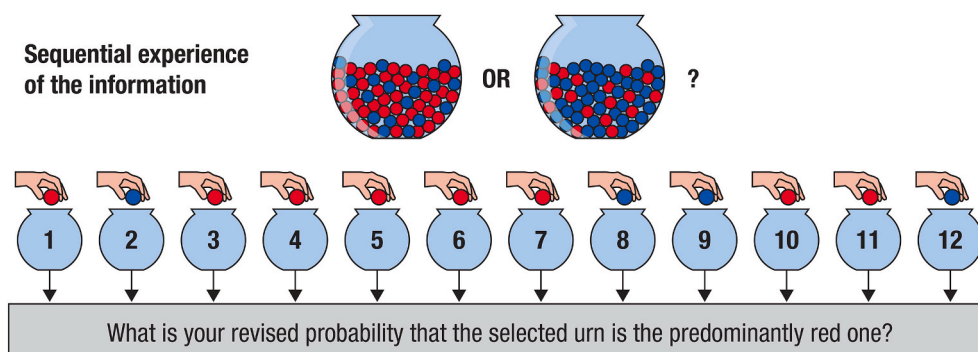

Sequential experience of the information

OR ?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |

What is your revised probability that the selected urn is the predominantly red one?

Fig. 2. Illustrative comparison of paradigmatic description- and experience-based Bayesian inference problems. In the description-based task (upper panel), participants read a summary of an inference problem and return a single probability judgment about the aggregate evidence (adapted from Kahneman & Tversky, 1972). In the experience-based task (lower panel), participants are asked to imagine two populations and then sequentially experience the actual presentation of evidence from one, randomly selected, population and return a revised probability judgment after each new observation (see, e.g., Edwards et al., 1965). In both example scenarios, the solution prescribed by Bayes's theorem is 97%.

protocols in work under the intuitive statistician umbrella. Moreover, they showed that this shift in experimental culture continues to have a profound influence on empirical research on probabilistic reasoning and human rationality.

2.3. Of Smart Animals and "Irrational" Humans

A final inconsistency in research on statistical intuitions concerns the extent to which probabilistic reasoning competences (or the lack thereof) are uniquely human. The origins of statistical abilities have, in fact, been traced beyond human ontogeny. Several nonhuman primates, including great apes, capuchin monkeys, and rhesus macaques, have been shown to draw flexible statistical inferences from populations to samples (De Petrillo & Rosati, 2019; Eckert, Call, Hermes, Herrmann, & Rakoczy, 2018; Rakoczy et al., 2014; Tecwyn et al., 2017; but see Placì, Eckert, Rakoczy, & Fischer, 2018) and, to a more limited extent, from samples to populations (Eckert, Rakoczy, & Call, 2017). Likewise, the ability to make probabilistic inferences has been demonstrated in different bird species (Bastos & Taylor, 2020; Clements, Gray, Gross, & Pepperberg, 2018) and some birds integrate physical information about objects when making statistical inferences (Bastos & Taylor, 2020). Finally, like babies, both chimpanzees and a bird species have been shown to integrate information about the psychological states of the sampling agents (humans), such as expressed preferences or visual access, when drawing statistical inferences (Bastos & Taylor, 2020; Eckert, Rakoczy, et al., 2018). It seems that "[i]ntuitive statistics in its most basic form is thus an evolutionarily more ancient rather than a uniquely human capacity" (Rakoczy et al., 2014, p. 60). Table 3 summarizes representative examples of this work on animals' statistical intuitions and describes the experimental procedure used. Because most animals, unlike humans, are not able to process and produce symbolic representations, animals' statistical intuitions and studies thereof are based on experience rather than description.

In addition, studies investigating animal decision making have shown that the decisions of several nonhuman species, unlike those of human adults, conform to the axioms of rational choice theory in some contexts (e.g., Edwards & Pratt, 2009; Monteiro, Vasconcelos, & Kacelnik, 2013; Rivalan, Winter, & Nachev, 2017; Schuck-Paim & Kacelnik, 2002). Moreover, animals show normatively correct behavior in situations involving sunk costs (Arkes & Ayton, 1999) and they often appear to be surprisingly "risk savvy" (e.g., Gallistel, 1990; Hanus & Call, 2014; Kheifets & Gallistel, 2012; Real, 1991; Real & Caraco, 1986; Weber et al., 2004). Again, we suggest that one key to resolving this apparent paradox lies in the degree of experiential involvement in the tasks used (see also Oaksford & Hall, 2016; Weber et al., 2004).

Assuming that humans' ability to cope with an uncertain world has evolved from animals' capacities to face volatile, competitive, and hard-to-predict environments, we might expect human and animal behavior to converge when humans are tested in situations in which they, too, make decisions from experience rather than description. Indeed, it has been found that when animals and humans accumulate information via direct experience, their risk sensitivity is predicted by a measure of risk per unit of return—the coefficient of variation—rather than by outcome variance, which is typically used in normative economic models (Weber et al., 2004). Similarly, the divergence between humans' and other animals' tendency to show a "certainty effect" in repeated risky choice could be resolved by taking into account differences in the experiential aspect of perceptual precision between the respective paradigms (Shafir, Reich, Tsur, Erev, & Lotem, 2008). Whereas animals typically experience primary rewards whose exact quantity is difficult to discern when consumed (e.g., water), humans often receive monetary rewards that can be precisely differentiated. When the perceptual noise in the presentation of rewards was increased, humans showed the same pattern of choices as animals; when it was decreased, honeybees displayed the same choice pattern as humans (Shafir et al., 2008). Finally, when making risky foraging decisions, "bumblebees underperceive rare events

Table 3

Summary of Representative Experience-Based Procedures Used in Studies Testing Animals' Statistical Intuitions.

Paradigm and study	Animal subjects	Sampling materials	Experience-based procedure and measurement	Physical instantiation
Single sample from population				
Bastos and Taylor (2020)	Kea	Different ratios of rewarding and unrewarding tokens in two transparent jars. In Exp. 2, a physical barrier divided the sampling population, impeding sampling from part of the population	One object was drawn from each population and animals made a single forced choice between the samples drawn while they were still concealed from view. In Exp. 3, whether or not sampling was random was varied (with or without visual access to the population during sampling) and the two samples were drawn by different experimenters	Direct interaction
Clements et al. (2018)	Grey parrot	3:1 ratio of two different types of objects (e.g., 3 corks and 1 piece of paper), visibly placed in an opaque bucket	One object was randomly drawn from the population. The animal vocally identified the object drawn while it was still concealed from view	Direct interaction
De Petrillo and Rosati (2019)	Rhesus macaques	Different ratios of artificial fruit (e.g., apples and lemons) inside a transparent lottery machine with one opening	The objects were shuffled inside the lottery machine until one exited through the opening; VoE	Distanced demonstration
Eckert, Call, et al. (2018)	Chimpanzees	Different ratios of favorable (e.g., peanuts, banana pellets, grapes) and neutral (e.g., monkey chow, carrot pieces) food items in two transparent containers	One food item was drawn from each population and animals made a single forced choice between the samples drawn while they were still concealed from view. In Eckert, Rakoczy, et al. (2018) whether or not sampling was random was varied (with or without visual access to the population during sampling) and the two samples were drawn by different experimenters	Direct interaction
Eckert, Rakoczy, et al. (2018)	Chimpanzees			
Placì et al. (2018)	Long-tailed macaques			
Rakoczy et al. (2014)	Great apes			
Tecwyn et al. (2017)	Capuchin monkeys			
Repeated sampling from population				
Eckert et al. (2017)	Great apes	Different ratios of favorable (fruit pellets) and neutral (carrot pieces) food items in two transparent containers	Two populations were shown to animals, then occluded, and the positions shuffled. Next, multiple food items were simultaneously drawn from each population in one action (Exp. 1: five items from each population; Exp. 2: three or five vs. 12 items from each population). Animals made a single forced choice between the occluded populations	Direct interaction

Note. VoE: Violation-of-expectation looking paradigm.

and overperceive common events" (Real, 1991, p. 985), which is the reverse of the inverse S-shaped probability weighting observed in description-based risky choices in humans (Kahneman & Tversky, 1979). When humans, like bees, rely on sampled experience, their weighting pattern also appears to reverse (when conditioned on objective probabilities; Regenwetter & Robinson, 2017) or to indicate more linear weighting (Wulff et al., 2018).

Thus, when compared on equal terms, animals, infants, and adults in research conducted in the 1960s and 1970s no longer seem so far apart in terms of the quality of their probabilistic reasoning.

3. Empirical Support for a Description–Experience Gap in Statistical Intuitions

This experience-based perspective on research on statistical intuitions gives rise to a strong, testable prediction, namely, that variations in the information format that increase or decrease the level of experiential involvement will lead, ceteris paribus, to corresponding changes in performance. This point implies that we, like others (Rakow & Newell, 2010), understand description and experience not in terms of a strict dichotomy but as concepts spread along a continuum.

On the descriptive side of the continuum, symbolic descriptions that reflect more of the underlying experience should lead to better statistical inferences. For example, described frequency formats have been argued to represent a proxy of the original experience (i.e., a series of events) on which cognitive algorithms have evolved (Cosmides & Tooby, 1996; Gigerenzer, 1991; Gigerenzer & Hoffrage, 1995). Described *natural frequencies* are defined as the product of natural sampling, which is the process of encountering observations of raw rather than normalized counts of events, without the marginal frequencies being fixed a priori (Kleiter, 1994). As such, natural frequencies maintain information about the underlying sample size and base rates, which is lost in conditional probabilities (Gigerenzer & Hoffrage, 1995). There is substantial evidence that descriptive formats harnessing natural frequencies improve Bayesian inference. A recent meta-analysis, reviewing 20 years of research on the "natural frequency facilitation effect" in Bayesian inference, estimated that the proportion of participants who correctly solved Bayesian inference problems presented in a natural frequency format was 20 percentage points higher than that solving problems presented as conditional probabilities (McDowell & Jacobs, 2017). Presenting information in a frequency format also improves other types of probabilistic reasoning—for example, in conjunction rule tasks (Fiedler, 1988; Hertwig & Gigerenzer, 1999; Tversky & Kahneman, 1983), sample size tasks (Sedlmeier & Gigerenzer, 1997), and the Monty Hall problem (Krauss & Wang, 2003). Importantly, although natural frequencies reflect some of the underlying experience, they remain a descriptive format. They can thus not be used to study probabilistic reasoning in babies and nonhuman primates, which requires truly experience-based methodologies—and thus moving further away from the descriptive side of the continuum.

For example, we would expect adding first-hand experience to reasoning tasks to further improve statistical inference. Indeed, allowing adults to directly experience simulated outcomes of probabilistic processes in several otherwise description-based inference tasks (e.g., conjunction rule tasks, the maternity ward problem, Bayesian inference, and the Monty Hall problem) drastically improved judgments (Hogarth & Soyer, 2011). Moreover, research on medical judgments has found that Bayesian inferences are more accurate when based on experiential formats (e.g., a sequence of representative patient cases illustrating the relative frequency of a disease and of positive/negative test results) than on described probability formats (Armstrong & Spaniol, 2017; Wegier & Shaffer, 2017). Similarly, a seminal review of base-rate fallacy research concluded that "when base rates are directly experienced through trial-by-trial outcome feedback, their impact on judgments increases" (Koehler, 1996, p. 6), in contrast to the relative "neglect" of stated base rates. Finally, people hold accurate distributional priors about the

duration and extent of everyday phenomena—such as human lifespans or movie run-time lengths—which can be learned from everyday experience with the statistical texture of real-world environments (Griffiths & Tenenbaum, 2006).

The work reviewed thus far has investigated the role of experience either within one subject population—typically human adults—or within a single task format across different subject populations. But to what extent can the description–experience continuum cast light on inconsistencies across subject populations and research traditions? This question is to some degree theoretical. Direct comparisons of all subject populations' statistical intuitions across the entire spectrum of the description–experience continuum are not feasible; systematic comparisons, where they are possible, are scarce. There are, however, important exceptions. For instance, in a study examining children's and adults' Bayesian probability updating in a natural frequency format versus a conditional probability format, sixth graders' Bayesian inferences in the natural frequency format matched those of adults in the conditional probability format (Zhu & Gigerenzer, 2006; but see Pighin, Girotto, & Tentori, 2017). Similarly, in a comparison of children's and adults' statistical intuitions on experience- or description-based paradigmatic probabilistic inference tasks—conjunction rule tasks and Bayesian inference tasks—we recently showed that sequentially experiencing statistical information considerably improved both adults' and children's statistical intuitions (Schulze & Hertwig, 2021). In fact, adults' probabilistic reasoning performance in description was surpassed by that of children in experience—seemingly a developmental reversal that was, we argue, in reality driven by the experimental protocol.

In sum, there is considerable empirical evidence that the description–experience dimension plays a key role when it comes to interpreting probabilistic reasoning within and across the lifespan, research traditions, and species. But the experimental paradigms we have reviewed of course also differ on other dimensions, and our view does not exclude the possibility of genuine evolutionary or developmental differences in probabilistic inference. Next, we summarize further factors that may play a role in determining the statistical competencies of different subject populations.

4. Other Possible Contributors to Inconsistencies in Research on Statistical Intuitions

4.1. Do Experience-Based Tasks Measure More Basic Statistical Skills and Are They Thus Easier?

Perhaps the most important alternative explanation for the inconsistencies observed is that studies with infants and animals target basic statistical abilities that adults possess as a matter of course. Yet the empirical findings we have reviewed suggest that description–experience gaps in performance cannot be attributed solely to a difficult–easy task dichotomy. When participants of the *same age* were given more experience- or more description-based problems, their statistical inferences differed systematically, even when the abstract statistical principle (e.g., Bayes's rule) remained the same (Armstrong & Spaniol, 2017; Fiedler, 1988; Hertwig & Gigerenzer, 1999; Hogarth & Soyer, 2011; McDowell & Jacobs, 2017; Wegier & Shaffer, 2017). Importantly, when participants of *different ages* were given inference problems that used different information formats but measured the same underlying statistical ability (e.g., adherence to the conjunction rule), performance differed depending on the information format (Schulze & Hertwig, 2021; Zhu & Gigerenzer, 2006). These findings suggest that experience-based paradigms do not produce better results solely because the statistical abilities engaged are more basic, even though sometimes this may be the case. Moreover, differences in task difficulty are not easily defined. If anything, one might expect the experience-based protocols used in the intuitive statistician tradition to be cognitively more taxing—in terms of memory, attention, and computation—than

equivalent description-based formats in which all information is packaged and delivered at once (Lejarraga & Hertwig, 2021). Finally, the risk of casting differences in performance between described and experienced experimental protocols as differences in task difficulty is that it merely re-describes the empirical data, unless the cognitive processes potentially underlying such a reduction in difficulty are specified (see 5.1. Mechanisms Underlying the Beneficial Qualities of Experience).

4.2. Implicit Versus Explicit Measures?

Another possible contributor to performance differences resides in the nature of the dependent measure used. Whereas infants' probabilistic understanding is often inferred from relatively implicit measures such as looking time (see Table 1), adults are typically asked for explicit verbal (written) judgments (see Table 2). Explicit responses may engage cognitive functions not needed for implicit responses (e.g., inhibitory control, working memory, or attention), thus contributing to the inconsistencies observed. The distinction between implicit and explicit measures has proven central in accounting for age-related discrepancies in other cognitive domains, such as false-belief understanding (Onishi & Baillargeon, 2005; Scott & Baillargeon, 2017) and language processing (Creel & Quam, 2015), and has also been examined in the context of probabilistic reasoning (e.g., Téglás et al., 2007; Yost, Siegel, & Andrews, 1962). Yet, again, several findings cannot be readily explained by this distinction. For instance, almost all studies probing animals' statistical intuitions (see Table 3; and see some studies with human infants, e.g., Denison & Xu, 2010b, 2014) required the animal to make an explicit, albeit nonverbal, decision. Research conducted with adults in the intuitive statistician and heuristics-and-biases traditions also tended to use explicit dependent measures.

4.3. Genuine Developmental or Evolutionary Differences?

Our perspective does not imply that infants' remarkable statistical intuitions attest to a cognitive competence that adults, for unclear reasons, no longer possess. Rather, we generally expect adults to equal or exceed infants' and children's performance when the task format is held constant (see Schulze & Hertwig, 2021), unless there are good empirical or theoretical reasons to the contrary. This view does not preclude the possibility that infants' probabilistic reasoning is genuinely different from adults', or that younger learners sometimes outperform older ones. For instance, it has been argued that humans' conceptual repertoire in dealing with numerical content shows considerable *discontinuities* across development (Carey, 2009) or that children's representation of probabilistic information may be profoundly different from that of adults, inasmuch as the role of heuristic cognitive processes appears to increase rather than decrease as a function of life experience (Reyna & Brainerd, 1995). Moreover, young children sometimes outperform older children or adults within description-based or experience-based paradigms (Arkes & Ayton, 1999; Davidson, 1995; Gopnik, Griffiths, & Lucas, 2015; Gualtieri & Denison, 2018), and statistical intuitions demonstrated in infants have not been confirmed in 3- and 4-year old children (Girotto, Fontanari, Gonzalez, Vallortigara, & Blaye, 2016; Girotto & Gonzalez, 2008). Counterintuitively, some of these discontinuities and even reversals may be the price for growth in other cognitive abilities: Biased responses could supplant unbiased ones because the knowledge structures needed for the operation of a specific heuristic (e.g., fully formed social stereotypes) are established at a later age (Davidson, 1995; Jacobs & Klaczynski, 2002; Jacobs & Potenza, 1991; Stanovich, West, & Toplak, 2011).

Another view rooted in evolutionary developmental psychology is that the greater cognitive constraints under which infants or animals operate—more generally, cognitive immaturity—can enable adaptive cognitive functioning (Bjorklund, 1997). For instance, restrictions in memory span may foster language acquisition by initially focusing the learner on simple grammatical relationships, laying the foundation for

the later acquisition of more complex grammatical structures (Elman, 1993; Newport, 1990). Similarly, prolonged prefrontal immaturity in early development delays the onset of stringent cognitive control, which may help children to learn social and linguistic conventions (Thompson-Schill, Ramscar, & Chrysikou, 2009) or to engage in broader internal and external exploration behaviors (Gopnik, 2020; Gopnik et al., 2015). Finally, animals' cognitively simpler architecture may enable them to better adhere to axioms of rational choice because they encode less complex contextual or symbolic information (Stanovich, 2013) and are less likely to abstract and overgeneralize learned rules (Arkes & Ayton, 1999).

4.4. Smart Babies, Smart Adults?

Finally, several lines of research on adult cognition paint a more optimistic picture of adults' competence to deal with probabilistic information than research in the heuristics-and-biases tradition does. Probabilistic models of cognition assume that human cognition at all ages can be explained in terms of a rational Bayesian framework, casting new light on several core cognitive functions (Clark, 2013; Friston, 2010; Oaksford & Chater, 2001; Tenenbaum, Kemp, Griffiths, & Goodman, 2011) and the course of cognitive development (Bonawitz, Denison, Griffiths, & Gopnik, 2014; Gopnik, 2012; Gopnik & Wellman, 2012; Perfors, Tenenbaum, Griffiths, & Xu, 2011; Xu, 2019). The related notion of resource rationality suggests that human cognition is guided by the optimal use of limited computational resources, meaning a rational trade-off between the costs and benefits of using computationally sophisticated analytic versus more heuristic strategies (Griffiths, Lieder, & Goodman, 2015; Lieder, Griffiths, & Hsu, 2018). More generally, the concept of bounded and ecological rationality has highlighted that—provided that people's cognitive strategies fit the structure of their environment—even simple strategies can lead to accurate decisions (Gigerenzer, Todd, & the ABC Research Group, 1999; Hertwig, Pleskac, Pachur, & the Center for Adaptive Rationality, 2019; Simon, 1955, 1956).

In summary, the ability to reason probabilistically is likely determined by various features of the inference task at hand as well as by the specific capacities and characteristics of the person facing the task. One key feature that has not yet been thoroughly considered in the literature is the experiential involvement in a task. Next, we highlight important properties of experience that distinguish it from description and identify mechanisms by which they can, under some circumstances, foster accurate judgments and decisions.

5. What Characterizes Experience and When Is It Beneficial?

Direct interaction with the world affords myriad concurrent dimensions of information that symbolic description lacks or conveys in condensed form, if at all (Hertwig et al., 2018). A learner experiencing a sequence of events may concurrently receive sensory and motoric feedback; obtain temporal, structural (e.g., clustering), and sample size information; or gain first-hand insights into conditions for statistical inferences (e.g., randomness) that need to be explicitly stipulated or assumed in descriptions. But what are the specific mechanisms by which first-hand experience facilitates appropriate statistical inference, and under which conditions can properties of experience improve or impair accurate statistical intuitions? Several underlying mechanisms are conceivable (see also Lejarraga & Hertwig, 2021) and it is likely that different factors determine the influence of experience on statistical intuitions, depending on the task at hand.

5.1. Mechanisms Underlying the Beneficial Qualities of Experience

5.1.1. Computational ease
One mechanism by which the sequential experience of events can reduce difficulty and improve judgments is by easing computational

demands. For instance, in choices between *n*-armed bandits, a simple method for estimating an action's value is to sum up the sampled rewards and weigh the sum by the sample size (Sutton & Barto, 1998). In experience-based choice, decision makers can thus maximize expected returns (based on the sampled experience) with a much simpler calculus than description-based expected value theory, which requires probabilities to be estimated and multiplied by rewards (Hertwig & Pleskac, 2010). In a recent meta-analysis of research on description- and experience-based choices between risky lotteries (using the sampling paradigm), a median of 55% of decisions from description maximized expected value, whereas 66% of decisions from experience maximized the experienced mean returns when the sampling sequence included all possible outcomes, including the rare one, relative to 89% when it did not (Wulff et al., 2018). Moreover, ongoing updating processes can alleviate the burden of storing all experienced outcomes (Frey, Mata, & Hertwig, 2015); experiential approaches may therefore become more important as decision problems become more complex (Hogarth & Soyer, 2015; Lejarraga, 2010).

Computational ease has also been proposed to explain why descriptions that retain some properties of the underlying experience, such as natural frequencies, improve reasoning. Natural frequencies have been argued to render Bayesian inferences computationally simpler because they preserve information about base rates, whereas conditional probability formats require this information to be incorporated into the calculation via additional computational steps (Gigerenzer & Hoffrage, 1995; McDowell & Jacobs, 2017). Thus, although different representation formats may be informationally equivalent, they may not trigger computationally equivalent cognitive algorithms (Gigerenzer & Hoffrage, 1995). Moreover, abstract symbolic information formats can introduce sources of ambiguity that experience-based information resolves more easily. In descriptive conjunction rule problems, for instance, people have been found to infer nonmathematical meanings of the term "probability" (Hertwig & Gigerenzer, 1999). Frequency formats of the same task disambiguate its semantic interpretation by highlighting the mathematical context and the applicability of the conjunction rule (Hertwig & Gigerenzer, 1999).

Finally, experience often affords control over both the source and the amount of information that needs to be processed and people can adjust their cognitive strategies based on their goals, cognitive abilities, and past experience. This aspect of experience-based reasoning can change the problem landscape such that the prevalence of more distinct, and thus easier, choice situations varies as a function of experiential involvement (Hertwig & Pleskac, 2010; Wulff et al., 2018). When decision problems are equated by matching task difficulty across description and experience—e.g., when comparing performance on experience-based perceptuo-motor tasks with that on equivalent described monetary gambles—there may be little difference in performance (Jarvstad et al., 2013; Wu et al., 2009). In many cases, however, it may not be possible to precisely quantify and thus equate computational ease. This would require an algorithmic implementation of the cognitive processes engaged by experience- and description-based tasks within a broader cognitive architecture, which would in turn need to consider developmental changes in the cognitive processes engaged by different task formats (see 5.2.2. Experience and description across the lifespan). One challenge for future research is to map properties of description- and experience-based tasks onto cognitive functions.

5.1.2. Incremental learning

Another core quality of experience is that it often involves repeated trials, permitting people to adapt to the environment and its demands, and thereby learn to exploit its affordances, in a step-by-step manner. Heuristics-and-biases research has been criticized for focusing on one-shot situations, thereby lacking such a continuous perspective and excluding "one of the most important determinants of the behavior they purport to explain" (Hogarth, 1981, p. 213). The distinction between repeated trials and one-shot studies is also key to understanding

sometimes surprisingly disparate experimental practices in economics and parts of psychology (Hertwig & Ortmann, 2001). One-shot studies leave little room for people to learn—to make incremental changes in light of new information, explore diverse solution pathways, correct behavioral consequences, and reach solutions even without being able to explain how and why (see, e.g., learning direction theory; Selten, 1998). Metaphorically speaking, sequential experience offers decision makers the opportunity to incrementally clip a hedge, whereas one-shot descriptive situations expect them to fell a tree in a single pass (cf. Connolly, 1988).

5.1.3. Internal states as a source of information

Finally, direct interaction with the environment affects the experiencing organism in a variety of ways, and the accumulation of these influences can serve as a source of information: "any physiological or psychological state variable that is altered by experience might function as an efficient integrator (a 'memory') of past experiences" (Higginson, Fawcett, Houston, & McNamara, 2018, p. 8). In a foraging context, reliance on a readily accessible physiological state, such as an animal's energy reserve, has been shown to be nearly as effective for survival as an optimal Bayesian learning strategy, which explicitly integrates all encountered experiences (Higginson et al., 2018). Similarly, psychological indicators such as emotions or moods can help to adaptively adjust behavior under environmentally uncertain conditions (Nettle & Bateson, 2012). Experience thus affords a means to exploit the structure of the environment by tapping into information about external conditions stored in an organism's internal states.

5.2. Experience Is Not a Panacea: Under Which Conditions Is It Beneficial?

5.2.1. Wicked environments and meta-cognitive myopia

Although experience entails properties that can improve statistical inference, it is no guarantee for good probabilistic reasoning. One important determinant of the value of first-hand experience is the conditions under which it is gathered (Erev & Roth, 2014; Hogarth, 2001). If an environment is "kind," in that it maps experience onto valid representations, intuitive statistics can be remarkably accurate; if an environment is "wicked," offering experience that is unrepresentative, biased, or misleading, statistical intuitions may be mistaken (Hogarth, 2001; Hogarth, Lejarraga, & Soyer, 2015). For instance, there is ample evidence that people automatically encode accurate frequency information about events they encounter in the world (Hasher & Zacks, 1979, 1984; Zacks & Hasher, 2002). Yet people sometimes behave as if they were held hostage by their experience—unwilling or unprepared to go beyond the data, and, in a kind of *meta-cognitive myopia,* failing to take the history and validity of the sample and the sampling process into account (Fiedler, 2000, 2012). These findings have led some researchers to describe humans as "*naïve* intuitive statisticians" (Fiedler & Juslin, 2006, emphasis added; but see Le Mens & Denrell, 2011, for a challenge to this naivety conclusion). Moreover, research on the description–experience gap in risky choice has shown that people do not always perform better when choosing from experience, although their choices in the sampling paradigm (see Hertwig & Erev, 2009) tend to be well described as maximization of mean return based on sampled experience (Hertwig & Pleskac, 2010; Wulff et al., 2018). Furthermore, in experimental paradigms featuring sequences of choices with feedback (e.g., the full feedback paradigm; Hertwig & Erev, 2009), systematic deviations from maximization have been observed—often opposite to those obtained in description (Erev et al., 2017; Plonsky et al., 2015). Animals sometimes likewise make suboptimal choices or violate principles of rational choice (e.g., Bateson, Healy, & Hurly, 2002; Shafir et al., 2008; Shafir, Waite, & Smith, 2002; Zentall, 2015).

Finally, description-based formats also have a place in people's information ecologies. It may not always be feasible to rely on direct experience of information (e.g., when facing novel realities) and there

are many situations in which descriptions offer clear benefits over experience (e.g., precision in distinguishing minute differences; summary representations of other people's experience). In many situations, moreover, people have access to first-hand experience in addition to a symbolic description. Drawing on research investigating the interplay between described and experienced information can thus advance our understanding of people's probabilistic thinking—for instance, when anticipating how people will perceive and respond to real-world risks, both existing and novel (Hertwig & Wulff, 2021).

5.2.2. Experience and description across the lifespan

The role of experiential involvement may also be complicated by developmental changes in the cognitive processes engaged by experience- and description-based tasks. For instance, experiential formats can require costly information acquisition that taxes memory and fluid intelligence. These demands may contribute to developmental differences in decision making (Rakow & Rahim, 2010). One example of a complex experiential task is the Iowa gambling task, in which adults typically outperform children (see Boyer, 2006). Giving children prior information about the options' outcomes and probabilities in addition to online experience of the outcomes (thus decreasing search and memory demands) has been shown to improve their performance (van Duijvenvoorde, Jansen, Bredman, & Huizenga, 2012). The cognitive demands of experiential formats may also affect decision making in older age. In complex choice environments (e.g., with numerous options), older adults (mean age = 71 years) explored substantially less than younger adults (mean age = 24 years) did, and their lower sampling effort was correlated with declines in measures of fluid intelligence (Frey et al., 2015). Moreover, there may be age-related differences in the affective processes triggered by immediate outcome experience, causing adolescents to take more risks than adults in, for instance, description-based dynamic choice tasks with immediate outcome feedback (Figner, Mackinlay, Wilkening, & Weber, 2009).

Likewise, the ability to deal with described probability formats may undergo significant change during development and may potentially require formal instruction through schooling. Research has shown that, when asked to choose between lotteries with stated outcomes and probabilities, children between 5 and 7 years old do not systematically consider differences in expected value between choice options (Levin, Weller, Pederson, & Harshman, 2007). Similarly, children younger than 9 years old do not systematically use stated probabilistic cues to inform their decisions in information-board paradigms; rather, they rely on unsystematic or unsuitable strategies (Betsch & Lang, 2013). In sum, making proportional calculations on the basis of described probabilistic information appears to be a difficult skill to learn (see also Bryant & Nunes, 2012). One challenge for future research is to further map out the developmental trajectory of the relationship between experience- and description-based probabilistic reasoning.

5.2.3. Cognitive illusions in experience-based paradigms

Finally, although the cognitive illusions reported in the context of heuristics-and-biases research predominantly occur in description-based tasks, there are important exceptions. One is the *gambler's fallacy*, which describes the belief that, in a random sequence such as the flip of a coin, a long streak of one outcome (heads) is more likely to be followed by the opposite outcome (tails) than would be expected under random sampling. The gambler's fallacy has been argued to arise from the misconception of chance processes as locally representative: "people expect that the essential characteristics of the process will be represented, not only globally in the entire sequence, but also locally in each of its parts" (Tversky & Kahneman, 1974, p. 1125). An alternative view suggests that the gambler's fallacy, and people's seeming misperception of randomness more generally, occur precisely because of how people experience statistical environments: Human experience is finite, and short-term memory capacity is limited (Hahn & Warren, 2009). In a finite sequence of 20 coin tosses, for instance, people may only be able to

monitor a limited number of events (e.g., four consecutive tosses) as their attentional window moves through this data stream. Under these conditions, there is a good chance that they will never encounter a sample of four consecutive heads; in fact, this probability is more than twice that of never seeing three heads followed by one tail (Hahn & Warren, 2009). That is, under conditions that match people's typical experience with finite samples, a sequence associated with committing the gambler's fallacy (three heads followed by one tail) is much more likely to occur than a sequence of four consecutive heads.

The notion of a "belief in the law of small numbers" (Tversky & Kahneman, 1971) has also been invoked to explain a classic choice anomaly that occurs when people make repeated choices from experience. For example, when asked to repeatedly choose between two alternatives with unequal odds of the same payoff (e.g., $p = .70$ and $1 - p = .30$), people tend to *probability match* by allocating responses to the choice options in proportion to their relative rates of success (Newell & Schulze, 2017; Vulkan, 2000). Again, this phenomenon may be driven by people's typical experience with repeated choices among probabilistic options. That is, probability matching in laboratory experiments may be an overlearned response from common real-world settings in which it can be a highly successful strategy due to competition for available resources (Gallistel, 1990; Schulze, van Ravenzwaaij, & Newell, 2015) or sequential dependencies in the outcome sequence (Gaissmaier & Schooler, 2008; Schulze, Gaissmaier, & Newell, 2020; Schulze, van Ravenzwaaij, & Newell, 2017). Thus, given the structure and affordances of the environments people typically encounter in their daily lives, paradigmatic experience-based "cognitive illusions" no longer seem so fallacious. Moreover, for some experience-based phenomena that have been labelled fallacious, such as the belief in a *hot hand* (Gilovich, Vallone, & Tversky, 1985), the bias may in fact reside in how researchers evaluate the adequacy of people's probabilistic reasoning (see Miller & Sanjurjo, 2014, 2018).

6. Implications of the Description–Experience Continuum of Statistical Intuitions

The ability to make good statistical inferences is a hallmark prerequisite for coping with the demands of an uncertain world. To navigate uncertainty successfully, humans and other animals must be able to draw apposite inferences from finite samples. This ability is likely determined both by properties of the inference task and by the capacities and characteristics of the individual facing it. One key aspect that has not yet been thoroughly considered is how decision makers acquire the data on which their inferences are based: via experience, symbolic description, or something in between. The description–experience framework we have outlined can help to explain the inconsistencies between infant and adult probabilistic reasoning and other contradictory findings in the young history of behavioral decision research. Let us conclude by outlining four key implications of the description–experience continuum of statistical intuitions: one methodological, one normative, one conceptual, and one relating to policy.

Concerning methodology, we have shown that diverse paradigms are needed to fully understand the breadth and adaptability of human probabilistic thinking. Experimental approaches that reduce people's statistical intuitions to a snapshot of their shortcomings run the danger of dismissing people (adults) as fundamentally inept in matters of probabilistic thinking—a prevalent and influential portrayal of the intuitive statistical mind in research on adult judgment and decision making that has been accentuated by current research on infant and animal probabilistic reasoning (see Gopnik, 2014). More generally, as our review highlights, researchers' choice of experimental design is crucial for determining the level of statistical competency they will observe. If an experimental task does not accurately represent the situation toward which a researcher intends to generalize, the processes studied may be altered in such a way that the results obtained are no longer representative of people's functioning in the natural causal

texture of their ecology (Brunswik, 1955; Dhami, Hertwig, & Hoffrage, 2004). Our argument is not that experience-based situations are more representative of people's information ecology than description-based situations—although they may arguably be more prevalent—or that they better represent the situations toward which researchers aim to generalize. Rather, we emphasize that both classes of situations exist in the information ecology of the 21st century and that it is important to strive to represent and contrast both designs wherever possible when investigating people's statistical intuitions. Only then will it be possible to gain a comprehensive understanding of the intuitive statistical competences of the human mind.

Second, the description–experience continuum of statistical intuitions has normative implications. Continuous experimental conditions require the evaluation of continuous learning processes (Hogarth, 1981). In description-based studies, researchers commonly assume benchmarks of statistical reasoning that hold in large samples or at the limit. Yet these principles do not always hold in small or finite samples—which arguably represent a more veridical reflection of the environmental statistics faced by real-world people with limited attention and short-term memory. It follows that "evaluations of rationality based on long-run considerations or limit properties may lead to a distorted picture" of human rationality (Hahn, 2014, p. 239). One task for future research will be to test experimental demonstrations of classic cognitive illusions not only against norms that hold at the limit but also against the reality of humans' limited experience (see Hahn, 2014; Hogarth, 1981).

A third implication is conceptual and points to directions for future research. We have suggested that the level of experiential involvement is crucial in determining the availability of various cognitive strategies that may differ in computational complexity, and that experience-based formats can enable the use of computationally simpler algorithms (see Gigerenzer & Hoffrage, 1995; Hertwig & Pleskac, 2010). Future research will need to further investigate the cognitive mechanisms by which experience improves or impairs probabilistic reasoning, and to determine what it is about symbolic descriptions that can make them difficult. Is it that they subtract "analogical" properties of the original experience (e.g., sequential experience in time)? Or that symbols are cultural inventions that humans have not evolved to process intuitively? Do they introduce sources of ambiguity that experience resolves more easily (e.g., semantic ambiguity; Hertwig, Benz, & Krauss, 2008; Hertwig & Gigerenzer, 1999)? And to what extent can descriptions, because they carry the author's implicit frame of reference (McKenzie & Nelson, 2003), make people more susceptible to being manipulated or misled? Of course, there will also be important contexts in which symbolic descriptions of the probabilistic texture of the world are indispensable because they, for instance, represent and summarize the accumulated wisdom and experience of others. Here the question is how such symbolic descriptions can be constructed to ensure maximum transparency and accessibility.

A related avenue for future research is to embed the description–experience distinction into research on the developmental trajectory of how probability is computed in the mind. Recent advances in research on probabilistic models of cognitive development suggest that human cognition is guided by powerful rational learning mechanisms from infancy onward (e.g., Bonawitz et al., 2014; Gopnik, 2012; Gopnik & Wellman, 2012; Perfors et al., 2011; Xu, 2019). Infants' remarkable statistical inference abilities have been interpreted as evidence for this theoretical framework because they indicate that the vital tools needed for rational statistical learning are already available in early infancy.

Finally, we turn to potential educational and other policy implications. Our review has been informed by a growing number of experiments on risky decisions from experience versus description (see, e.g., Wulff et al., 2018). This research has highlighted implications for policy in the world outside the laboratory: in communicating risk, designing economic markets, and implementing safe practices in the workplace and everyday life (Barron, Leider, & Stack, 2008; Erev & Roth, 2014; Kaufmann, Weber, & Haisley, 2013; Weber, 2006; Yechiam, Erev, &

Barron, 2006). Extending this conceptual distinction to developmental issues promises to generate fruitful research questions and, ultimately, educational applications. Some of the implications of the description–experience continuum that we have highlighted, such as the use of natural frequency-based formats to foster statistical inference, have been discussed in medical, legal, and educational settings (see, e.g., Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000). Another avenue for future research is to better understand the relationship between intuitive, experience-based notions of probability and the more formal views of probability acquired through schooling. A wealth of developmental research suggests that making proportional calculations of probability is a difficult skill to learn (for a review, see Bryant & Nunes, 2012). But might it be conceivable to capitalize on infants' well-formed intuitions when teaching children to solve formal probability problems? Gopnik (2014) suggests that it may indeed be possible to "exploit [babies'] intuitive abilities to teach children, and adults, to understand probability better and to make better decisions as a result." We have shown that one crucial aspect in pursuing this goal is to take into account experiential features that can foster good statistical intuitions. Finally, it is possible that enlisting experience in the process of teaching and learning probabilistic reasoning—for instance, by experiencing the processes of information sampling or the consequences of cognitive acts (e.g., inferences, choices, and estimates), by engaging with and systematically manipulating physical instantiations of chance devices, or by being exposed to stationary and nonstationary environments in which the information sampled and the decisions made entail different degrees of contingency (that are complicated to describe symbolically)—is better suited to empower self-directed and active learning (see Gureckis & Markant, 2012) than is the processing of symbolic information.

Let us conclude by returning to Gopnik's (2014) question. Our review of the large and disparate literature on statistical intuitions suggests that one key to understanding the puzzling discrepancy between smart babies and stupid grown-ups in matters of probabilistic reasoning is that infants and children typically operate on the basis of immediate experience or require information formats that closely approximate that experience. Adults—and this is one of the great achievements of cognitive development—are also able to process symbolic, propositional representations of the world. Their untutored statistical intuitions may, however, be better tuned to experience- than to description-based formats.

CRediT authorship contribution statement

Christin Schulze: Conceptualization, Visualization, Writing - original draft, Writing - review & editing. **Ralph Hertwig**: Conceptualization, Writing - review & editing.

Declaration of Competing Interest

None.

Acknowledgments

References

Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin, 125*(5), 591–600. https://doi.org/10.1037/0033-2909.125.5.591.

Armstrong, B., & Spaniol, J. (2017). Experienced probabilities increase understanding of diagnostic test results in younger and older adults. *Medical Decision Making, 37*(6), 670–679. https://doi.org/10.1177/0272989X17691954.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063.

Attisano, E., & Denison, S. (2020). Infants' reasoning about samples generated by intentional versus non-intentional agents. *Infancy, 25*(1), 110–124. https://doi.org/10.1111/infa.12320.

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance, 9*(3), 396–406. https://doi.org/10.1016/0030-5073(73)90061-5.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16*(3), 215–233. https://doi.org/10.1002/bdm.443.

Barron, G., Leider, S., & Stack, J. (2008). The effect of safe experience on a warnings' impact: Sex, drugs, and rock-n-roll. *Organizational Behavior and Human Decision Processes, 106*(2), 125–142. https://doi.org/10.1016/j.obhdp.2007.11.002.

Bastos, A. P. M., & Taylor, A. H. (2020). Kea show three signatures of domain-general statistical inference. *Nature Communications, 11*, Article 828. https://doi.org/10.1038/s41467-020-14695-1.

Bateson, M., Healy, S. D., & Hurly, T. A. (2002). Irrational choices in hummingbird foraging behaviour. *Animal Behaviour, 63*(3), 587–596. https://doi.org/10.1006/anbe.2001.1925.

Betsch, T., & Lang, A. (2013). Utilization of probabilistic cues in the presence of irrelevant information: A comparison of risky choice in children and adults. *Journal of Experimental Child Psychology, 115*(1), 108–125. https://doi.org/10.1016/j.jecp.2012.11.003.

Bjorklund, D. F. (1997). The role of immaturity in human development. *Psychological Bulletin, 122*(2), 153–169. https://doi.org/10.1037/0033-2909.122.2.153.

Bonawitz, E., Denison, S., Griffiths, T. L., & Gopnik, A. (2014). Probabilistic models, learning algorithms, and response variability: Sampling in cognitive development. *Trends in Cognitive Sciences, 18*(10), 497–500. https://doi.org/10.1016/j.tics.2014.06.006.

Boyer, T. W. (2006). The development of risk-taking: A multi-perspective review. *Developmental Review, 26*(3), 291–345. https://doi.org/10.1016/j.dr.2006.05.002.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*(3), 193–217. https://doi.org/10.1037/h0047470.

Bryant, P., & Nunes, T. (2012). *Children's understanding of probability: A literature review*. Nuffield Foundation.

Camilleri, A. R., & Newell, B. R. (2019). Better calibration when predicting from experience (rather than description). *Organizational Behavior and Human Decision Processes, 150*, 62–82. https://doi.org/10.1016/j.obhdp.2018.10.006.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477.

Clements, K. A., Gray, S. L., Gross, B., & Pepperberg, I. M. (2018). Initial evidence for probabilistic reasoning in a grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology, 132*(2), 166–177. https://doi.org/10.1037/com0000106.

Connolly, T. (1988). Hedge-clipping, tree-felling, and the management of ambiguity: The need for new images of decision-making. In L. R. Pondy, R. J. Boland, Jr., & H. Thomas (Eds.), *Managing ambiguity and change* (pp. 37–50). Wiley.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*(1), 1–73. https://doi.org/10.1016/0010-0277(95)00664-8.

Creel, S. C., & Quam, C. (2015). Apples and oranges: Developmental discontinuities in spoken-language processing? *Trends in Cognitive Sciences, 19*(12), 713–716. https://doi.org/10.1016/j.tics.2015.09.006.

Dai, J., Pachur, T., Pleskac, T. J., & Hertwig, R. (2019). What the future holds and when: A description–experience gap in intertemporal choice. *Psychological Science, 30*(8), 1218–1233. https://doi.org/10.1177/0956797619858969.

Davidson, D. (1995). The representativeness heuristic and the conjunction fallacy effect in children's decision making. *Merrill-Palmer Quarterly, 41*(3), 328–346.

De Petrillo, F., & Rosati, A. G. (2019). Rhesus macaques use probabilities to predict future events. *Evolution and Human Behavior, 40*(5), 436–446. https://doi.org/10.1016/j.evolhumbehav.2019.05.006.

Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology, 49*(2), 243–249. https://doi.org/10.1037/a0028278.

Denison, S., Trikutam, P., & Xu, F. (2014). Probability versus representativeness in infancy: Can infants use naïve physics to adjust population base rates in probabilistic inference? *Developmental Psychology, 50*(8), 2009–2019. https://doi.org/10.1037/a0037158.

Denison, S., & Xu, F. (2010a). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science, 34*(5), 885–908. https://doi.org/10.1111/j.1551-6709.2010.01111.x.

Denison, S., & Xu, F. (2010b). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science, 13*(5), 798–803. https://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*(3), 335–347. https://doi.org/10.1016/j.cognition.2013.12.001.

Denison, S., & Xu, F. (2019). Infant statisticians: The origins of reasoning under uncertainty. *Perspectives on Psychological Science, 14*(4), 499–509. https://doi.org/10.1177/1745691619847201.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin, 130*(6), 959–988. https://doi.org/10.1037/0033-2909.130.6.959.

Dutt, V., Arló-Costa, H., Helzner, J., & Gonzalez, C. (2014). The description–experience gap in risky and ambiguous gambles. *Journal of Behavioral Decision Making, 27*(4), 316–327. https://doi.org/10.1002/bdm.1808.

Eckert, J., Call, J., Hermes, J., Herrmann, E., & Rakoczy, H. (2018). Intuitive statistical inferences in chimpanzees and humans follow Weber's law. *Cognition, 180*, 99–107. https://doi.org/10.1016/j.cognition.2018.07.004.

Eckert, J., Rakoczy, H., & Call, J. (2017). Are great apes able to reason from multi-item samples to populations of food items? *American Journal of Primatology, 79*(10), Article e22693. https://doi.org/10.1002/ajp.22693.

Eckert, J., Rakoczy, H., Call, J., Herrmann, E., & Hanus, D. (2018). Chimpanzees consider humans' psychological states when drawing statistical inferences. *Current Biology, 28*(12), 1959–1963. https://doi.org/10.1016/j.cub.2018.04.077.

Edwards, S. C., & Pratt, S. C. (2009). Rationality in collective decision-making by ant colonies. *Proceedings of the Royal Society B: Biological Sciences, 276*(1673), 3655–3661. https://doi.org/10.1098/rspb.2009.0981.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). Wiley.

Edwards, W., Lindman, H., & Phillips, L. D. (1965). Emerging technologies for making decisions. In *New Directions in Psychology II* (pp. 261–325). Holt, Rinehart and Winston.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*(1), 71–99. https://doi.org/10.1016/0010-0277(93)90058-4.

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review, 124*(4), 369–409. https://doi.org/10.1037/rev0000062.

Erev, I., & Roth, A. E. (2014). Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences, 111*(Supplement 3), 10818–10825. https://doi.org/10.1073/pnas.1402846111.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50*(2), 123–129. https://doi.org/10.1007/BF00309212.

Fiedler, K. (2000). Beware of samples! A cognitive–ecological sampling approach to judgment biases. *Psychological Review, 107*(4), 659–676. https://doi.org/10.1037/0033-295X.107.4.659.

Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical inference. In B. H. Ross (Ed.), The psychology of learning and motivation (Vol. 57, pp. 1–55). Academic Press. doi:10.1016/B978-0-12-394293-7.00001-7.

Fiedler, K., & Juslin, P. (2006). Taking the interface between mind and environment seriously. In K. Fiedler, & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 3–29). Cambridge University Press.

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 709–730. https://doi.org/10.1037/a0014983.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*(24), 15822–15826. https://doi.org/10.1073/pnas.232472899.

Fontanari, L., Gonzalez, M., Vallortigara, G., & Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. *Proceedings of the National Academy of Sciences, 111*(48), 17075–17080. https://doi.org/10.1073/pnas.1410583111.

Frey, R., Mata, R., & Hertwig, R. (2015). The role of cognitive abilities in decisions from experience: Age differences emerge as a function of choice set size. *Cognition, 142*, 60–80. https://doi.org/10.1016/j.cognition.2015.05.004.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. https://doi.org/10.1038/nrn2787.

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition, 109*(3), 416–422. https://doi.org/10.1016/j.cognition.2008.09.007.

Gallistel, C. R. (1990). *The organization of learning*. MIT Press.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology, 2*(1), 83–115. https://doi.org/10.1080/14792779143000033.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*(3), 295–314. https://doi.org/10.1016/0010-0285(85)90010-6.

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. https://doi.org/10.1016/j.cognition.2016.03.010.

Girotto, V., & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition, 106*(1), 325–344. https://doi.org/10.1016/j.cognition.2007.02.005.

Glaser, C., Trommershäuser, J., Mamassian, P., & Maloney, L. T. (2012). Comparison of the distortion of probability information in decision under risk and an equivalent

visual task. *Psychological Science, 23*(4), 419–426. https://doi.org/10.1177/0956797611429798.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research and policy implications. *Science, 337*(6102), 1623–1627. https://doi.org/10.1126/science.1223416.

Gopnik, A. (2014). The surprising probability gurus wearing diapers. *The Wall Street Journal*. Retrieved from http://www.wsj.com/articles/SB10001424052702303393804579308662389246416.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 375*(1803), Article 20190502. https://doi.org/10.1098/rstb.2019.0502.

Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science, 24*(2), 87–92. https://doi.org/10.1177/0963721414556653.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*(6), 1085–1108. https://doi.org/10.1037/a0028044.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science, 7*(2), 217–229. https://doi.org/10.1111/tops.12142.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17*(9), 767–773. https://doi.org/10.1111/j.1467-9280.2006.01780.x.

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

Güney, Ş., & Newell, B. R. (2015). Overcoming ambiguity aversion through experience. *Journal of Behavioral Decision Making, 28*(2), 188–199. https://doi.org/10.1002/bdm.1840.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science, 7*(5), 464–481. https://doi.org/10.1177/1745691612454304.

Gweon, H., & Schulz, L. E. (2011). 16-month-olds rationally infer causes of failed actions. *Science, 332*(6037), 1524. https://doi.org/10.1126/science.1204493.

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences, 107*(20), 9066–9071. https://doi.org/10.1073/pnas.1003095107.

Hahn, U. (2014). Experiential limitation in judgment and decision. *Topics in Cognitive Science, 6*(2), 229–244. https://doi.org/10.1111/tops.12083.

Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review, 116*(2), 454–461. https://doi.org/10.1037/a0015241.

Hanus, D., & Call, J. (2014). When maths trumps logic: Probabilistic judgements in chimpanzees. *Biology Letters, 10*(12), Article 20140892. https://doi.org/10.1098/rsbl.2014.0892.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*(3), 356–388. https://doi.org/10.1037/0096-3445.108.3.356.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*(12), 1372–1388. https://doi.org/10.1037/0003-066X.39.12.1372.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x.

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of *and. Cognition, 108*(3), 740–753. https://doi.org/10.1016/j.cognition.2008.06.008.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences, 13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004.

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12*(4), 275–305. https://doi.org/10.1002/(SICI)1099-0771(199912)12:4<275::AID-BDM323>3.0.CO;2-M.

Hertwig, R., Hogarth, R. M., & Lejarraga, T. (2018). Experience and description: Exploring two paths to knowledge. *Current Directions in Psychological Science, 27*(2), 123–128. https://doi.org/10.1177/0963721417740645.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences, 24*(3), 383–451.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition, 115*(2), 225–237. https://doi.org/10.1016/j.cognition.2009.12.009.

Hertwig, R., Pleskac, T. J., Pachur, T., & the Center for Adaptive Rationality. (2019). Taming uncertainty. MIT Press.

Hertwig, R., & Wulff, D. (2021). A description–experience framework in psychology of risk. Manuscript submitted for publication.

Higginson, A. D., Fawcett, T. W., Houston, A. I., & McNamara, J. M. (2018). Trust your gut: Using physiological states as a source of information is almost as effective as optimal Bayesian learning. *Proceedings of the Royal Society B: Biological Sciences, 285*, Article 20172411. https://doi.org/10.1098/rspb.2017.2411.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290*(5500), 2261–2262. https://doi.org/10.1126/science.290.5500.2261.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin, 90*(2), 197–217. https://doi.org/10.1037/0033-2909.90.2.197.

Hogarth, R. M. (2001). *Educating intuition.* The University of Chicago Press.

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science, 24*(5), 379–385. https://doi.org/10.1177/0963721415591878.

Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus nontransparent description. *Journal of Experimental Psychology: General, 140*(3), 434–463. https://doi.org/10.1037/a0023265.

Hogarth, R. M., & Soyer, E. (2015). Providing information for decision making: Contrasting description and simulation. *Journal of Applied Research in Memory and Cognition, 4*(3), 221–228. https://doi.org/10.1016/j.jarmac.2014.01.005.

Jacobs, J. E., & Klaczynski, P. A. (2002). The development of judgment and decision making during childhood and adolescence. *Current Directions in Psychological Science, 11*(4), 145–149. https://doi.org/10.1111/1467-8721.00188.

Jacobs, J. E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development, 62*(1), 166–178. https://doi.org/10.2307/1130712.

Jarvstad, A., Hahn, U., Rushton, S. K., & Warren, P. A. (2013). Perceptuo-motor, cognitive, and description-based decision-making seem equally good. *Proceedings of the National Academy of Sciences, 110*(40), 16271–16276. https://doi.org/10.1073/pnas.1300239110.

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237–251. https://doi.org/10.1037/h0034747.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291. https://doi.org/10.2307/1914185.

Kaufmann, C., Weber, M., & Haisley, E. (2013). The role of experience sampling and graphical displays on one's investment risk appetite. *Management Science, 59*(2), 323–340. https://doi.org/10.1287/mnsc.1120.1607.

Kayhan, E., Gredebäck, G., & Lindskog, M. (2018). Infants distinguish between two events based on their relative likelihood. *Child Development, 89*(6), e507–e519. https://doi.org/10.1111/cdev.12970.

Kheifets, A., & Gallistel, C. R. (2012). Mice take calculated risks. *Proceedings of the National Academy of Sciences, 109*(22), 8776–8779. https://doi.org/10.1073/pnas.1205131109.

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. https://doi.org/10.1016/S0010-0277(02)00004-5.

Kleiter, G. D. (1994). Natural sampling: Rationality without base rates. In G. H. Fischer, & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375–388). Springer.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*(1), 1–53. https://doi.org/10.1017/S0140525X00041157.

Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall problem: Discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General, 132*(1), 3–22. https://doi.org/10.1037/0096-3445.132.1.3.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science, 21*(8), 1134–1140. https://doi.org/10.1177/0956797610376652.

Lawson, C. A., & Rakison, D. H. (2013). Expectations about single event probabilities in the first year of life: The influence of perceptual and statistical information. *Infancy, 18*(6), 961–982. https://doi.org/10.1111/infa.12014.

Le Mens, G., & Denrell, J. (2011). Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. *Psychological Review, 118*(2), 379–392. https://doi.org/10.1037/a0023010.

Lejarraga, T. (2010). When experience is better than description: Time delays and complexity. *Journal of Behavioral Decision Making, 23*(1), 100–116. https://doi.org/10.1002/bdm.666.

Lejarraga, T., & Hertwig, R. (2021). How experimental culture shapes views of human rationality. Manuscript submitted for publication.

Lejarraga, T., & Lejarraga, J. (2020). Confidence and the description–experience distinction. *Organizational Behavior and Human Decision Processes, 161*, 201–212. https://doi.org/10.1016/j.obhdp.2020.07.003.

Lejarraga, T., Pachur, T., Frey, R., & Hertwig, R. (2016). Decisions from experience: From monetary to medical gambles. *Journal of Behavioral Decision Making, 29*(1), 67–77. https://doi.org/10.1002/bdm.1877.

Lejarraga, T., Woike, J. K., & Hertwig, R. (2016). Description and experience: How experimental investors learn about booms and busts affects their financial risk taking. *Cognition, 157*, 365–383. https://doi.org/10.1016/j.cognition.2016.10.001.

Levin, I. P., Weller, J. A., Pederson, A. A., & Harshman, L. A. (2007). Age-related differences in adaptive decision making: Sensitivity to expected value in risky choice. *Judgment and Decision Making, 2*(4), 225–233.

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review, 125*(1), 1–32. https://doi.org/10.1037/rev0000074.

Ma, L., & Xu, F. (2011). Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition, 120*(3), 403–411. https://doi.org/10.1016/j.cognition.2011.02.003.

Ma, L., & Xu, F. (2013). Preverbal infants infer intentional agents from the perception of regularity. *Developmental Psychology, 49*(7), 1330–1337. https://doi.org/10.1037/a0029620.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77–80. https://doi.org/10.1126/science.283.5398.77.

Martin, J. M., Gonzalez, C., Juvina, I., & Lebiere, C. (2014). A description–experience gap in social interactions: Information about interdependence and its effects on cooperation. *Journal of Behavioral Decision Making, 27*(4), 349–362. https://doi.org/10.1002/bdm.1810.

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin, 143*(12), 1273–1312. https://doi.org/10.1037/bul0000126.

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review, 10*(3), 596–602. https://doi.org/10.3758/BF03196520.

Miller, J. B., & Sanjurjo, A. (2014). A cold shower for the hot hand fallacy. In *IGIER Working Paper No. 518*. https://doi.org/10.2139/ssrn.2450479.

Miller, J. B., & Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica, 86*(6), 2019–2047. https://doi.org/10.3982/ECTA14943.

Monteiro, T., Vasconcelos, M., & Kacelnik, A. (2013). Starlings uphold principles of economic rationality for delay and probability of reward. *Proceedings of the Royal Society B: Biological Sciences, 280*(1756), Article 20122386. https://doi.org/10.1098/rspb.2012.2386.

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological Science, 21*(7), 960–969. https://doi.org/10.1177/0956797610372637.

Nettle, D., & Bateson, M. (2012). The evolutionary origins of mood and its disorders. *Current Biology, 22*(17), R712–R721. https://doi.org/10.1016/j.cub.2012.06.020.

Newell, B. R., & Schulze, C. (2017). Probability matching. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory* (2nd ed., pp. 62–78). Routledge.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*(1), 11–28. https://doi.org/10.1207/s15516709cog1401_2.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5*(8), 349–357. https://doi.org/10.1016/S1364-6613(00)01699-5.

Oaksford, M., & Hall, S. (2016). On the source of human irrationality. *Trends in Cognitive Sciences, 20*(5), 336–344. https://doi.org/10.1016/j.tics.2016.03.002.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258. https://doi.org/10.1126/science.1107621.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120*(3), 302–321. https://doi.org/10.1016/j.cognition.2010.11.015.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68*(1), 29–46. https://doi.org/10.1037/h0024722.

Pighin, S., Girotto, V., & Tentori, K. (2017). Children's quantitative Bayesian inferences from natural frequencies and number of chances. *Cognition, 168*, 164–175. https://doi.org/10.1016/j.cognition.2017.06.028.

Placì, S., Eckert, J., Rakoczy, H., & Fischer, J. (2018). Long-tailed macaques (*Macaca fascicularis*) can use simple heuristics but fail at drawing statistical inferences from populations to samples. *Royal Society Open Science, 5*(9), Article 181025. https://doi.org/10.1098/rsos.181025.

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review, 122*(4), 621–647. https://doi.org/10.1037/a0039413.

Rakoczy, H., Clüver, A., Saucke, L., Stoffregen, N., Gräbener, A., Migura, J., & Call, J. (2014). Apes are intuitive statisticians. *Cognition, 131*(1), 60–68. https://doi.org/10.1016/j.cognition.2013.12.011.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*(1), 1–14. https://doi.org/10.1002/bdm.681.

Rakow, T., & Rahim, S. B. (2010). Developmental insights into experience-based decision making. *Journal of Behavioral Decision Making, 23*(1), 69–82. https://doi.org/10.1002/bdm.672.

Real, L. A. (1991). Animal choice behavior and the evolution of cognitive architecture. *Science, 253*(5023), 980–986. https://doi.org/10.1126/science.1887231.

Real, L. A., & Caraco, T. (1986). Risk and foraging in stochastic environments. *Annual Review of Ecology and Systematics, 17*, 371–390. https://doi.org/10.1146/annurev.es.17.110186.002103.

Regenwetter, M., & Robinson, M. M. (2017). The construct–behavior gap in behavioral decision research: A challenge beyond replicability. *Psychological Review, 124*(5), 533–550. https://doi.org/10.1037/rev0000067.

Rehder, B., & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & Cognition, 45*(2), 245–260. https://doi.org/10.3758/s13421-016-0662-3.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*(1), 1–75. https://doi.org/10.1016/1041-6080(95)90031-4.

Rivalan, M., Winter, Y., & Nachev, V. (2017). Principles of economic rationality in mice. *Scientific Reports, 7*, Article 17441. https://doi.org/10.1038/s41598-017-17747-7.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4.

Saffran, J. R., Reeck, K., Niebuhr, A., & Wilson, D. (2005). Changing the tune: The structure of the input affects infants' use of absolute and relative pitch. *Developmental Science, 8*(1), 1–7. https://doi.org/10.1111/j.1467-7687.2005.00387.x.

Schmandt-Besserat, D. (1996). *How writing came about.* University of Texas Press.

Schuck-Paim, C., & Kacelnik, A. (2002). Rationality in risk-sensitive foraging choices by starlings. *Animal Behaviour, 64*(6), 869–879. https://doi.org/10.1006/anbe.2003.2003.

Schulze, C., Gaissmaier, W., & Newell, B. R. (2020). Maximizing as satisficing: On pattern matching and probability maximizing in groups and individuals. *Cognition, 205*, Article 104382. https://doi.org/10.1016/j.cognition.2020.104382.

Schulze, C., & Hertwig, R. (2021). Statistical intuitions: Experiencing information improves children's and adults' probabilistic inferences. Manuscript submitted for publication.

Schulze, C., van Ravenzwaaij, D., & Newell, B. R. (2015). Of matchers and maximizers: How competition shapes choice under risk and uncertainty. *Cognitive Psychology, 78*, 78–98. https://doi.org/10.1016/j.cogpsych.2015.03.002.

Schulze, C., van Ravenzwaaij, D., & Newell, B. R. (2017). Hold it! The influence of lingering rewards on choice diversification and persistence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(11), 1752–1767. https://doi.org/10.1037/xlm0000407.

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences, 21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*(1), 33–51. https://doi.org/10.1002/(SICI)1099-0771(199703)10:1<33::AID-BDM244>3.0.CO;2-6.

Selten, R. (1998). Features of experimentally observed bounded rationality. *European Economic Review, 42*(3–5), 413–436. https://doi.org/10.1016/S0014-2921(97)00148-7.

Shafir, S., Reich, T., Tsur, E., Erev, I., & Lotem, A. (2008). Perceptual accuracy and conflicting effects of certainty on risk-taking behaviour. *Nature, 453*, 917–920. https://doi.org/10.1038/nature06841.

Shafir, S., Waite, T. A., & Smith, B. H. (2002). Context-dependent violations of rational choice in honeybees (Apis mellifera) and gray jays (Perisoreus canadensis). *Behavioral Ecology and Sociobiology, 51*(2), 180–187. https://doi.org/10.1007/s00265-001-0420-8.

Sim, Z. L., & Xu, F. (2017). Infants preferentially approach and explore the unexpected. *British Journal of Developmental Psychology, 35*(4), 596–608. https://doi.org/10.1111/bjdp.12198.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118. https://doi.org/10.2307/1884852.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129–138. https://doi.org/10.1037/h0042769.

Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice. *Thinking & Reasoning, 19*(1), 1–26. https://doi.org/10.1080/13546783.2012.713178.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31*(2–3), 103–118. https://doi.org/10.1016/j.dr.2011.07.003.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* MIT Press.

Tecwyn, E. C., Denison, S., Messer, E. J. E., & Buchsbaum, D. (2017). Intuitive probabilistic inference in capuchin monkeys. *Animal Cognition, 20*(2), 243–256. https://doi.org/10.1007/s10071-016-1043-9.

Téglás, E., & Bonatti, L. L. (2016). Infants anticipate probabilistic but not deterministic outcomes. *Cognition, 157*, 227–236. https://doi.org/10.1016/j.cognition.2016.09.003.

Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences, 104*(48), 19156–19159. https://doi.org/10.1073/pnas.0700271104.

Téglás, E., Ibanez-Lillo, A., Costa, A., & Bonatti, L. L. (2015). Numerical representations and intuitions of probabilities at 12 months. *Developmental Science, 18*(2), 183–193. https://doi.org/10.1111/desc.12196.

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science, 332*(6033), 1054–1059. https://doi.org/10.1126/science.1196404.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788.

Thompson-Schill, S. L., Ramscar, M., & Chrysikou, E. G. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science, 18*(5), 259–263. https://doi.org/10.1111/j.1467-8721.2009.01648.x.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences, 12*(8), 291–297. https://doi.org/10.1016/j.tics.2008.04.010.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105–110. https://doi.org/10.1037/h0031322.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458. https://doi.org/10.1126/science.7455683.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293.

van den Bos, W., & Hertwig, R. (2017). Adolescents display distinctive tolerance to ambiguity and to uncertainty during risky decision making. *Scientific Reports, 7*, Article 40962. https://doi.org/10.1038/srep40962.

van Duijvenvoorde, A. C. K., Jansen, B. R. J., Bredman, J. C., & Huizenga, H. M. (2012). Age-related changes in decision making: Comparing informed and noninformed situations. *Developmental Psychology, 48*(1), 192–203. https://doi.org/10.1037/a0025601.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys, 14*(1), 101–118. https://doi.org/10.1111/1467-6419.00106.

Weber, E. U. (2006). Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic Change, 77*, 103–120. https://doi.org/10.1007/s10584-006-9060-3.

Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review, 111*(2), 430–445. https://doi.org/10.1037/0033-295X.111.2.430.

Wegier, P., & Shaffer, V. A. (2017). Aiding risk information learning through simulated experience (ARISE): Using simulated outcomes to improve understanding of conditional probabilities in prenatal Down syndrome screening. *Patient Education and Counseling, 100*(10), 1882–1889. https://doi.org/10.1016/j.pec.2017.04.016.

Wellman, H. M., Kushnir, T., Xu, F., & Brink, K. A. (2016). Infants use statistical sampling to understand the psychological world. *Infancy, 21*(5), 668–676. https://doi.org/10.1111/infa.12131.

Wu, S.-W., Delgado, M. R., & Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences, 106*(15), 6088–6093. https://doi.org/10.1073/pnas.0900102106.

Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). Online product reviews and the description–experience gap. *Journal of Behavioral Decision Making, 28*(3), 214–223. https://doi.org/10.1002/bdm.1841.

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description–experience gap. *Psychological Bulletin, 144*(2), 140–176. https://doi.org/10.1037/bul0000115.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review, 126*(6), 841–864. https://doi.org/10.1037/rev0000153.

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition, 112*(1), 97–104. https://doi.org/10.1016/j.cognition.2009.04.006.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences, 105*(13), 5012–5015. https://doi.org/10.1073/pnas.0704450105.

Yechiam, E., Erev, I., & Barron, G. (2006). The effect of experience on using a safety device. *Safety Science, 44*(6), 515–522. https://doi.org/10.1016/j.ssci.2005.11.006.

Yost, P. A., Siegel, A. E., & Andrews, J. M. (1962). Nonverbal probability judgments by young children. *Child Development, 33*(4), 769–780. https://doi.org/10.2307/1126888.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & T. Betsch (Eds.), *Etc. frequency processing and cognition* (pp. 21–36). Oxford University Press.

Zentall, T. R. (2015). When animals misbehave: Analogs of human biases and suboptimal choice. *Behavioural Processes, 112*, 3–13. https://doi.org/10.1016/j.beproc.2014.08.001.

Zhang, H., & Houpt, J. W. (2020). Exaggerated prevalence effect with the explicit prevalence information: The description-experience gap in visual search. *Attention, Perception, & Psychophysics, 82*(7), 3340–3356. https://doi.org/10.3758/s13414-020-02045-8.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition, 98*(3), 287–308. https://doi.org/10.1016/j.cognition.2004.12.003.