The evolutionary roots of intuitive statistics

Dissertation

Zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

„Doktor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm Behavior and Cognition (BeCog)

der Georg-August University School of Science (GAUSS)

vorgelegt von

Johanna Eckert

aus Würzburg

Göttingen, 2018

Preliminary Note

The present thesis is a publication-based (cumulative) dissertation. It is based on three original articles that have been published:

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)*.* doi:10.1002/ajp.22693

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. doi: 10.1016/j.cognition.2018.07.004

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*. doi: 10.1016/j.cub.2018.04.077

In the present thesis, I relate the contents of the three articles to a broader theoretical context, summarize the main findings, and provide an extensive joint discussion. The original articles are reprinted in Appendices A, B, and C. I served as first author in all articles and was responsible for (a) developing the theory, (b) designing and conducting the experiments, (c) analyzing and interpreting the data, and (d) writing up and publishing the manuscripts. My thesis supervisors Hannes Rakoczy and Josep Call, as well as my co-authors supported me with helpful advice and discussion throughout the process.

I hereby declare that all parts of this dissertation were written by myself, assistance of third parties was only accepted if scientifically justifiable and acceptable in regards to the examination regulations, and all sources have been quoted.

Göttingen, 2018   *J.E.*

For Margarete

Thank you

To **Hannes Rakoczy** and **Josep Call** for choosing me to accomplish this project. I am extremely grateful for the opportunity to work with and learn from two such excellent researchers. Thank you for being a great supervisor-team, for trusting me and giving me all the freedom I wanted, while always offering me support and guidance where needed.

To **Julia Fischer**, who opened the door to the "primate world" for me five years ago. Thank you for further supporting me as a member of my thesis committee.

To **Tanya Behne, Annekathrin Schacht,** and **Michael Waldmann** for agreeing to be members of the examination board.

To **Esther Herrmann** for giving me the opportunity to work at Ngamba Island and to the **Leibniz-ScienceCampus Primate Cognition** for funding my research trips to Uganda.

To the **Ngamba Island Chimpanzee Sanctuary team**, especially **Innocent, Paul, Kyama, Byron, Joseph Kaale, Enos, Joseph Masereka**, **Amos**, and **Phillip** for their incredible effort to give us the best testing experience possible and to **Hategeka, Ronald,** and **Moses** for special culinary delights. A special thanks goes to the sunset cruise-team for creating some unforgettable memories.

To **Hanna Petschauer, Johannes Grossmann**, **Sebastian Schütte**, and all **zoo keepers** for enabling smooth and enjoyable data collection at the Leipzig Zoo.

To **Daniel Hanus** for a fabulous collaboration.

To **Jonas Hermes** for facilitating data collection for the human part of this dissertation.

To **Shona Duguid** for being a great island companion, for your support as study helper, for your always on-point scientific feedback and your encouragement and infinite gin supply in moments of doubt. It was a pleasure to be stuck on two hectares for several months with you.

To **Dairon Rodriguez** and **Stephan Kaufhold** for your highly appreciated help to chop the hundreds of thousands of food items needed for my studies. Stephan- your pellet-cutting-machine should be patented!

To **Sarah Placi, Michael Kriegl, Sandra Azar, Marina Proft, Dominik Eckert, Shona Duguid,** and **Antonia Bieber** for proof reading parts of an earlier draft of this thesis and **Bastian Lange** for help with formatting. Your comments and suggestions helped a lot to improve this manuscript.

To all my brilliant colleagues at the **MPI EVA** from whom I learned so much and who became good friends over the last years, most notably **Alex, Dairon, Daniel, Lou, Marina, Martin, Matthias, Sarah, Shona, Sonja, Steffi, Stephan, and Suska.**

Table of Contents

0. Abstract

Intuitive statistical reasoning is the capacity to draw intuitive probabilistic inferences based on an understanding of the relations between populations, sampling processes and resulting samples. This capacity is fundamental to our daily lives and one of the hallmarks of human thinking. We constantly use sample observations to draw general conclusions about the world, use these generalizations to predict what will happen next and to make rational decisions under uncertainty. Historically, statistical reasoning was thought to develop late in ontogeny, to be biased by general-purpose heuristics throughout adulthood, and to be restricted to certain situations and specific types of information. In the last decade, however, evidence has accumulated from developmental research showing that even pre-verbal infants can reason from populations of items to randomly drawn samples and vice versa. Moreover, infants can flexibly integrate knowledge from different cognitive domains (such as physical or psychological knowledge) into their statistical inferences. This indicates that neither language nor mathematical education are prerequisites for intuitive statistical abilities. Beyond that, recent comparative research suggests that basic forms of such capacities are not uniquely human: Rakoczy et al. (2014) presented nonhuman great apes with two populations with different proportions of preferred to non-preferred food items. Apes were able to infer which population was more likely to lead to a preferred food item as randomly drawn sample. Hence, just like human infants, great apes can reason from population to sample, giving a first hint that human statistical abilities may be based on an evolutionary ancient capacity.

The aim of the present dissertation is to explore the evolutionary roots of intuitive statistics more systematically and comprehensively by expanding on the initial findings of Rakoczy et al. (2014). I examined three questions regarding the i) generality and flexibility of nonhuman great apes´ statistical capacities, ii) their cognitive structures and limits, as well as iii) their interaction with knowledge from other cognitive domains. To address these questions, I conducted three studies applying variants of the paradigm established by Rakoczy et al. (2014).

In the first study, zoo-living great apes (*Pan troglodytes*, *Pan paniscus*, *Pongo abelii*, *Gorilla gorilla*) were required to infer from samples to populations of food items: Apes were presented with two covered populations and witnessed representative multi-item samples being drawn from these populations. Subsequently, subjects could choose which population they wanted to receive as a reward. I found that apes´ statistical abilities in this direction were more restricted than vice versa. However, these limitations were potentially due to accessory task demands rather than limitations in statistical reasoning. The second study was designed to gain deeper insights into the cognitive structure of intuitive statistics in

chimpanzees and humans. More specifically, I tested sanctuary-living chimpanzees and human adults in a task requiring inferences from population to sample and I systematically varied the magnitude of difference between the populations´ ratios (the ratio of ratios, ROR). I discovered that the statistical abilities of both chimpanzees and human adults varied as a function of the ROR and thus followed Weber´s law. This suggests that intuitive statistics are based on the analogue magnitude system, an evolutionary ancient cognitive mechanism common to many aspects of quantitative cognition. The third study investigated whether chimpanzees consider knowledge about others´ mental states when drawing statistical inferences. I tested sanctuary-living chimpanzees in a task that required subjects to infer which of two populations was more likely to lead to a desired outcome for the subject. I manipulated whether the experimenters had preferences to draw certain food types or acted neutrally and whether they had visual access to the populations while sampling or drew blindly. Chimpanzees chose based on proportional information alone when they had no information about experimenters' preferences and (to a lesser extent) when experimenters had preferences for certain food types but drew blindly. By contrast, when biased experimenters had visual access, subjects ignored statistical information and instead chose based on experimenters' preferences. Consistent with recent findings on pre-verbal infants, apes seem to have a random sampling assumption that can be overridden under the appropriate circumstances and they are able to use information about others´ mental states to judge whether this is necessary.

Taken together, the findings of the present dissertation indicate that nonhuman great apes possess intuitive statistical capacities on a par with those of human infants. Therefore, intuitive statistics antedate language and mathematical thinking not only ontogenetically, but also phylogenetically. This suggests that humans´ statistical abilities are founded on an evolutionary ancient capacity shared with our closest living relatives.

1. General introduction

1. General introduction

"Could there be in a normal man an intuition of probability just as fundamental and just as frequently used as, say, the intuition of whole numbers?"

(Piaget & Inhelder, 1975, xiii)

Some parts of life are consistent and entirely predictable. If I push my glass over the edge of the table, it will fall to the ground. If I go out without an umbrella while it rains, I will get wet. However, not all events happening around us follow such straightforward causal relations. In fact, virtually any situation involves some uncertain component. The glass will fall down when I push it, but will it break? I will get wet in case it rains, but what are the chances that the weather will change today? Our life is full of situations which require us to deal with variable, random, and only partly predictable features. We need to make judgments under uncertainty in the simplest of tasks, such as deciding whether to take an umbrella or not, and in more complex processes, for example when deciding whether to try a new migraine treatment or whether to invest in an insurance against storm damage. How are humans able to make such judgments and decisions?

The presumably optimal way to make decisions under uncertainty is to make probabilistic estimations. In many cases we can rely on our past experiences, perhaps updated with current information, to infer the most likely outcome of events. For example, before you call your friend, you may think about past times you tried to call her at this particular time and rapidly compare the proportion of times you reached her with the proportion of times she did not answer. This first estimation might be updated when you remember that she told you about a business meeting, resulting in low chances that she will answer her phone right now. In other cases, we are required to compute probabilities of events we have not experienced before by directly observing and judging proportional information. For instance, before you decide in which of two restaurants you go during your city trip, you may have a look at their ratings on the internet. If restaurant 1 has 100 positive ratings, but 200 bad ones, you may infer that restaurant 2, with just 20 positive, but only a single negative rating, will probably serve better food.

How do these probabilistic computations work? Do we have some cognitive mechanism sensitive to probabilistic input, allowing us to draw statistical inferences and to make rational predictions? As the above-mentioned statement taken from Piaget and Inhelder´s 1975 book exemplifies, the idea of an intuitive statistical inference mechanism has been fascinating researchers from various fields over many decades. However, despite the ubiquity of situations requiring statistical inferences, until today

psychological and cognitive research accumulated evidence showing that humans, children as well as adults, are surprisingly bad in tasks requiring explicit statistical inferences. This led to the conclusion that the ability to reason about probabilities develops late in human ontogeny and remains error-prone and tainted by heuristics and biases throughout adulthood.

It is only very recent that the establishment of non-verbal procedures allowed the development of appropriate experiments for pre- and non-verbal creatures, revealing astonishing results. Even pre-verbal infants have an intuitive sense of probability. They seem to understand the relation between populations (of events or objects), sampling processes (e.g. random, intentional, or otherwise determined) and resulting samples (individual events or objects) and are able to draw intuitive inferences regarding statistical matters. These findings indicate that humans in fact do possess a powerful domain-general statistical inference mechanism, which is independent of language and mathematical education and probably allows humans to rapidly learn about their environment from birth onwards.

These new insights raise a fundamental question concerning the phylogenetic origins of our intuitive statistical inference mechanism: Are we dealing with a uniquely human capacity, perhaps accounting for our unprecedented cognitive flexibility? Or do we share it with other animals, most notably our closest living relatives, the nonhuman great apes[1]? Comparative research investigating quantitative abilities found some striking similarities in basic numerical cognition of humans and nonhuman primates, which might give a first hint that great apes, too, possess the prerequisites for intuitive statistics. Even more importantly, recent research found that great apes are sensitive towards probabilistic information and can draw intuitive inferences from population to sample in much the same way as human infants do.

The aim of the present dissertation is to shed more light on the evolutionary roots of intuitive statistics by investigating this ability in great apes. In particular, I sought to get insights into the cognitive structure of intuitive statistics and, from a comparative point of view, to explore the commonalities and differences between great apes and humans. I was especially interested in three broad questions: First, how flexible are great apes´ statistical reasoning abilities - are they, like in human infants, bi-directional, i.e. from population to sample as much as from sample to population? Second, what are the cognitive structures and limits of intuitive statistics - are they the same in humans and in nonhuman great apes?

---

[1] In the following I will use the terms "great apes" and "apes" as synonyms to refer to the nonhuman great ape species.

And lastly, to which degree can nonhuman great apes integrate intuitive statistics with other types of information, a feature indicative of a domain-general inference mechanism?

In what follows, I will first introduce the most influential approaches to human statistical reasoning, as well as recent advances in developmental research (Section 2). Thereafter, I will provide a review of basic arithmetic abilities in nonhuman primates and their cognitive foundations (Section 3). Then, I will discuss the current body of research investigating probabilistic reasoning in nonhuman animals (Section 4) which will lay the foundation for the research questions of the present work (Section 5). Subsequently, I will present the three studies conducted in the course of this dissertation (Section 6), followed by a general discussion of the main findings in light of the evolutionary roots of intuitive statistics (Section 7).

2. Are humans intuitive statisticians?

The study of intuitive statistics, the capacity to draw intuitive probabilistic inferences based on an understanding of the relations between populations, sampling processes, and samples, has a long history in psychology. As long as this history of research has been, as diverse became the theories trying to explain the scope, functioning, and limits of statistical inferences. While Peterson and Beach (1967) have titled "Man as intuitive statistician", less than a decade later, Kahneman and Tversky´s seminal work concluded that "in making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction" (Kahneman & Tversky, 1973, p. 237). Yet another approach to human statistical reasoning stated that "the mind is an intuitive statistician of the frequentist school" (Gigerenzer, 1991, p. 9) and hence at least "some of our inductive reasoning mechanisms do embody aspects of calculus of probability" (Cosmides & Tooby, 1996, p. 3). The modern Bayesian approach to human statistical reasoning, on the contrary, views "inductive learning as a species of computational problems and the human mind as a natural computer" (Tenenbaum, Kemp, Griffiths, & Goodman, 2011, p. 1279). It was only relatively recently that developmental research revealed empirical evidence that humans indeed are intuitive statisticians from very early on: by the age of six months, humans already possess "the key prerequisite abilities for an inductive inference mechanism, based on the principles of rational Bayesian inference" (Denison & Xu, 2012, p. 51).

In the following I will first provide an overview over the most prominent theories of human statistical abilities- the heuristics and biases approach, the frequentist approach, and the Bayesian approach. Subsequently, I will review recent developmental work focusing on pre-verbal infants and discuss its implications for the various theories on human statistical reasoning.

2.1. The heuristics and biases approach

Research on how heuristics and biases shape human decision making has a long history in the social sciences. Even though others have worked on this topic already half a century earlier (e.g. Lippmann, 1922), with their seminal review "Judgment under uncertainty: Heuristics and biases" Amos Tversky and Daniel Kahneman (1974) became the luminaries of a steadily growing research discipline with tremendous impact on diverse fields such as law, medicine, economics, and management. This impact can be recognized by the fact that Kahneman was awarded the 2002 Nobel Prize in Economic Science. To this

day, their idea of intuitive probability judgment as suboptimal and rather irrational process that is heavily influenced by general-purpose heuristics is still widely esteemed.

In general, the idea of the heuristics and biases program is that, in order to assess probabilities and to make predictions, humans employ several types of heuristics (i.e. strategies, that ignore parts of the information), which reduce the complexity of probabilistic tasks to simpler judgmental operations. These mental shortcuts are quite efficient and extremely useful in everyday life, since they provide quick and effortless alternatives to slow and effortful statistical computations. In some cases, however, they can lead to systematic biases and errors. In the following I will give an overview about the two most popular types of heuristics employed in probabilistic tasks and the errors and biases they can cause. This overview will not be exhaustive, since it mainly focuses on the seminal work of Tversky and Kahneman. An overview about additional types of heuristics and biases can be found, e.g. in a review by Gigerenzer and Gaissmaier (2011).

2.1.1. Representativeness

The perhaps most famous type of heuristics is the representativeness heuristic, which is, according to Tversky and Kahneman (1974), used in tasks requiring inferences from population to sample and vice versa. In such tasks, people are usually asked to estimate the probability that a certain sample was drawn from or belongs to a particular population, or to judge the likelihood that a particular population will produce a certain sample. The representative heuristic enables humans to assess such probabilities by a simple rule of thumb: The more sample and population resemble and therefore are representative of each other, the higher the probability that the sample originated from the population. While this mental shortcut is most likely sufficient for many probabilistic problems, it disregards several key factors that should affect optimal judgments of probability, such as the base-rate probability and the sample size, and therefore can lead to systematic errors.

One such error is the conjunction fallacy, caused by a non-compliance of "perhaps the simplest and the most basic qualitative law of probability", the conjunction rule (Tversky & Kahneman, 1983, p. 293). The conjunction rule states that a conjunction cannot be more probable than each of its constituents. In other words, the probability of an event than includes two uncertain factors, cannot be higher than the probability of each of the two factors separately. The conjunction fallacy has classically been illustrated by the "Linda problem" (Lippmann, 1922). In this paradigm, participants are given a

character description of a fictive person (Linda) before they are asked to guess which of several occupations and/or avocations Linda has. When Linda is described with features stereotypical for academics and political activists, participants usually judge the likelihood that Linda is a bank teller *and* a feminist to be higher than Linda just being a bank teller (e.g. Tversky & Kahneman, 1983). This is presumably due to her personality description resembling the stereotypical description of a feminist, thereby seducing people to ignore the conjunction rule in favor of the representative heuristic. The conjunction fallacy has been and still is widely studied with humans of different age classes, varying levels of education, and in a wide range of contexts see, e.g. (see, e.g. Bonini, Tentori, & Osherson, 2004; Davidson, 1995; Fabre, Caverni, & Jungermann, 1995; Fantino, Kulik, Stolarz-fantino, & Wright, 1997; Fisk, 2005; Stanovich & West, 1998; Wells, 1985; Yates & Carlson, 1986). However, also a variety of alternative explanations have been offered. For example, it has been proposed that participants tested in the "Linda paradigm" may have misunderstood the word "and", or relatedly, that they interpreted the single statement to be exclusive of the other (e.g. that "Linda is a bank teller" excludes the possibility that Linda also is a feminist) and so correctly choose their conjunction (i.e. Linda is both a bank teller *and* a feminist (Chase, 1998; Dulany & Hilton, 1991; Gigerenzer, 1996; Hertwig & Gigerenzer, 1999; Macdonald & Gilhooly, 1990; Politzer & Noveck, 1991). Alternatively, participants may have assumed that the information provided by the experimenters must be relevant and therefore concluded that their answer should appreciate this relevant information (Chase, Hertwig, & Gigerenzer, 1998). While a range of studies have partly ruled out these explanations (see, e.g. Costello, 2009 for a review) the debate on how strong conjunction fallacy effects are in real life scenarios and under which exact circumstances they occur, is still ongoing (see, e.g. Charness, Karni, & Levin, 2010; Hertwig, Benz, & Krauss, 2008; Tentori & Crupi, 2012).

A similarly common fallacy presumably caused by the representativeness heuristic is base-rate neglect. This fallacy describes the phenomenon that humans often rely more on individuating information, i.e. information about a particular case, such as a personality description, instead of essential base-rate information to estimate probabilities. Kahneman and Tversky (1973), for instance, asked adults to judge whether a hypothetical person, allegedly sampled at random from a group of professionals, was rather an engineer or rather a lawyer, based on two pieces of information: A description of the base-rates of lawyers and engineers in the group (e.g. "70% of professionals are lawyers, 30% are engineers") and a brief personality description of the person in question (e.g. "he likes working on home carpentry projects"). Participants mainly judged based on the personality descriptions, widely ignoring the base-rate

probabilities. Hence, again people chose based on the degree to which the person in question resembled the classic stereotype of either lawyers or engineers. Interestingly, participants even neglected base-rate information when the provided personality description was completely irrelevant for the categorization. They only considered base-rates accurately when no information about personality traits was provided. Similar findings were obtained in studies replicating Kahneman and Tversky´s original study (1973), both in adults (e.g. Davidson & Hirtle, 1990; Ginosar & Trope, 1980; Manis, Dovalina, Avis, & Cardoze, 1980; Nisbett & Borgida, 1975; Pennycook, Fugelsang, & Koehler, 2012) and in children (e.g. Davidson, 1995; Jacobs & Potenza, 1991). Interestingly, base-rate neglect seems to be particularly apparent in social problems, compared to object domain problems, perhaps accounting for the emergence and durability of stereotypes (Jacobs & Potenza, 1991). Some studies demonstrated, however, that people can be sensitized to base-rate information under certain conditions (see, e.g. Erev, Shimonowitch, Schurr, & Hertwig, 2008 for discussion).

The representativeness heuristic can also lead to a general misconception of chance as a "self-correcting process". This misconception is the basis of recency effects, which describe the common, but wrong impression that a random event is more or less likely to occur, depending on whether it has or has not occurred in a while (Tversky & Kahneman, 1971). The negative recency effect (also known as gambler´s fallacy) is best illustrated in the following statement: "After observing a long run of red on the roulette wheel, for example, most people will erroneously believe that black is now due, presumably because the occurrence of black will result in a more representative sequence than the occurrence of an additional red" (Tversky & Kahneman, 1974, p. 1125). Conversely, people also exhibit positive recency effects (also known as hot hand fallacy), for example when they believe in having a lucky streak, and thus think an event is more likely to occur, because it has happened several times in a row (e.g. Gilovich, Vallone, & Tversky, 1985). Both types of fallacies can be observed in children and adults. Yet, the positive recency effect seems to decrease with age, whereas no such effect was found for the negative recency effect (Chiesi & Primi, 2009). The somewhat curious finding that the representative heuristic can lead to positive *and* negative recency effects (thus, the exact same data can cause an individual to expect two contrasting events) might be due to differences in people´s prior experiences (e.g. Ayton & Fischer, 2004).

Other important shortcomings of human intuitive statistics presumably induced by the representative heuristic are insensitivity to sample size (people judge based on whether or not the sampling proportion is representative of the population´s proportion, largely ignoring the size and thereby the meaningfulness of the sample; e.g. Kahneman & Tversky, 1972; Nisbett & Ross, 1980) and

misconceptions of regression (people do not consider regression to the mean and are surprised to see, e.g. an average performance after an outstanding performance in an event involving chance (Kahneman & Tversky, 1973)).

2.1.2. Availability

Another type of heuristic that is, according to Tversky & Kahneman (1974), frequently applied to judge the likelihood of an event, the numerosity of a class, or the frequency of co-occurrence of events, is the availability heuristic. This judgmental heuristic is based on the fact that "instances of large classes are usually recalled better and faster than instances of less frequent classes" (Tversky & Kahneman, 1974, p. 1127). Thus, people tend to assess the probability of an event by the ease with which occurrences of such an event can be remembered or imagined. For example, to assess the probability that a woman of a certain age class develops breast cancer, one may recall such instances within one´s own circle of acquaintances. As with the representative heuristic, the availability heuristic is probably a useful and efficient shortcut in many judgmental situations, but at the same time it is easily affected by factors other than frequencies and probabilities and therefore leads to predictable biases.

One such bias is due to the fact that the availability heuristic is strongly influenced by factors which do not reflect the actual probability of an event, but which alter the ease by which it is recalled. For example, events that are particularly salient, familiar, or recent are more easily recalled than others, irrespective of whether they are objectively likely to happen or not (Tversky & Kahneman, 1974). One may tend, for instance, to overestimate the objective risk of a middle-aged woman to be diagnosed with breast cancer, after one´s own sister was tested positive. Other factors like the imaginability of events (Galbraith & Underwood, 1973), or the perceived strength of association between two events (Chapman & Chapman, 1969) have similar effects.

Relatedly to various two-system theories of reasoning (e.g. Evans, 2008), the heuristics and biases program argues that intuitive judgments occupy a position "between the automatic parallel operations of perception, and the controlled serial operations of reasoning" (Kahneman & Frederick, 2002, p. 2). More specifically, the cognitive basis of probabilistic heuristics such as representativeness is supposedly the intuitive System 1, which guides quick, automatic, and associative operations. In contrast, System 2 is thought to be slow, self-aware, and controlled. In the heuristics and biases view, System 2 is monitoring the intuitive judgments of System 1 and has the power to override and correct them (see e.g. Kahneman & Frederick, 2002 for a review). Biases and fallacies are committed when System 2 fails to correct the flawed intuitions of System 1, for example due to time pressure or cognitive load (e.g. Finucane, Alhakami,

Slovic, & Johnson, 2000; Gilbert, 1989; Bodenhausen, 1990). The consequent conclusion of Tversky and Kahneman (1971, p. 31) was that intuitions (which are the source of heuristics) should always be regarded "with proper suspicion".

In sum, the heuristics and biases program has discovered extremely diverse and ubiquitous biases and errors that humans regularly commit when making decisions under uncertainty or when assessing the probability of an event: Humans do not sufficiently account for sample size, disregard the base-rate probability, ignore the conjunction rule, overestimate the prevalence of events that are recalled easily, and do not anticipate regression to the mean, just to name some of them. Hence, this line of research suggests that human intuitive statistical reasoning is fundamentally flawed, biased, and error prone. This view has not remained unchallenged. In particular, the interpretation of heuristics as error-prone intuitions and as source of irrationality has evoked considerable criticism. Gerd Gigerenzer, for instance, famously argued that heuristics, statistics and logic are each suited and adaptive for a particular type of problem (see, e.g. Gigerenzer & Gaissmaier, 2011 for a review): In real world scenarios, where parts of the relevant information are usually unknown, heuristics often depict the better option compared to statistical models, which require knowledge of all the relevant alternatives, their consequences, and probabilities (Simon, 1979). That this is at least partly true was demonstrated in empirical examples, where formalized heuristics led to better predictions than standard statistical methods which had the same ore more information, in diverse areas such as business, medical and legal decision making (see Gigerenzer & Gaissmaier, 2011 for a review). Does this mean humans are not as irrational as the heuristics and biases program indicated? A different approach on human statistical reasoning abilities argues that it all depends on the presentation format of the task. This important approach, the frequentist view, will be introduced in the next section.

2.2. The frequentist approach

Soon after the publication of Tversky and Kahneman´s seminal work, a new, alternative theory about human statistical reasoning arose with, in many regards, contrasting point of views. One essential aim of this new movement was to answer the following puzzle resulting from the heuristics and biases work: "If making accurate judgments under uncertainty is an important adaptive problem, why would natural selection have designed a mind that uses error-prone heuristics rather than an accurate calculus of probability?" (Cosmides & Tooby, 1996, p. 11).

The frequentist hypothesis (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995) states that "some of our inductive reasoning mechanisms do embody aspects of calculus of probability, but they are designed to take frequency information as input and produce frequency information as output" (Cosmides & Tooby, 1996, p. 3). More specifically, frequentists argue that humans and other animals regularly need to make accurate decisions under uncertainty in order to increase their chances of survival. Our ancestors, however, did not have access to explicit, i.e. symbolic forms of probabilistic information. Instead, they could only use their own experiences in form of encountered frequencies as database for probabilistic inferences. It makes thus sense, from an evolutionary point of view, that humans (and possibly other animals) evolved cognitive mechanisms that use exactly this observable information format as in- and output. Hence, frequentists concluded that humans have the capacity of inductive reasoning (i.e. to draw general conclusions from sample observations), as long as information is presented in frequency format (e.g. 1 out of 10 instances instead of 10%) and a frequency can be given as output (see Cosmides & Tooby, 1996).

This theory has two important implications: 1. Many of the findings of the heuristics and biases program might be explained by the fact that those tasks did not facilitate the representation of probabilities as frequencies. 2. Humans should not be able to reason about single-event probabilities. The latter can be explained as follows: Frequentists interpret probability as the relative frequency of an event defined over a specific reference class. A single-event, however, does not have a relative frequency - it either occurs or not. For example, there is no sense in stating that "tomorrow it will snow 30%" - either it will snow or not snow. Hence, according to frequentists, in realistic scenarios, single-case probabilities are meaningless. Moreover, a single-event does not have a single reference class. Instead, it can have an infinite number of reference classes. Cosmides and Tooby (1996) illustrated this with the following example: The relative frequency with which a woman aged between 35 and 40 gets breast cancer is 4 out of 100. What is the probability that Mrs. X, who is 49, and whose mother had breast cancer, will be diagnosed positive next year? According to Cosmides and Tooby it is completely meaningless to say the risk is "4%", because Mrs X could belong to an indefinitely large number of references classes (e.g. smoker, living a stressful life, mother of three, etc.) and the relative frequency of breast cancer may differ for each of them. Conversely, if one would figure in all possibly relevant characteristics, the number of reference classes would be limited, but then these classes would contain such a limited number of individuals that, due to an increased "error term", the relative frequency derived from these classes would again be completely meaningless. Therefore, Cosmides and Tooby (1996) conclude, "one cannot be a frequentist

and accept single-event probabilities". From an evolutionary point of view, as explained above, frequentists further argue that the human mind could not have adapted to detect single-event probabilities, because the necessary input format (i.e. probability expressed as numerical value) was only invented in modern times and could therefore not have been favored by natural selection.

Several studies aimed to test the predictions following from the frequentist hypotheses. For example, empirical tests have examined whether, as claimed by e.g. Gigerenzer (1991, p. 22), the conjunction fallacy and other "so-called cognitive illusions" disappear when the problem is phrased in frequentist way. While in some studies the rate of conjunction fallacies decreased, or the effect vanished completely (e.g. Fiedler, 1988; Hertwig & Gigerenzer, 1999), in others it remained stable when information was presented in form of frequencies (Kahneman & Tversky, 1996; Tentori, Bonini, & Osherson, 2004; Wedell & Moro, 2008). Even the trial of an adversarial collaboration project between a frequentist and one of the founders of the heuristics and biases program (Mellers, Hertwig, & Kahneman, 2001) did not end the controversy: Frequency formats did not reduce conjunction fallacies in the "Linda task", they, however, did so in a related task ("James task"). Other studies tested the influence of frequency formats on probabilistic reasoning in children. Zhu and Gigerenzer (2006), for example, found evidence that 9-to-12-year-old children could not reason correctly about probabilities when information was presented in terms of probabilities. When the same information was presented as natural frequencies, however, the proportion of children´s correct probability estimates increased. A more recent study aimed to replicate these findings (Pighin, Tentori, & Girotto, 2017) and could not find any evidence for natural frequency format being more easily processed or more likely to elucidate correct inferences than chance format. Hence, it is still debated whether and under which conditions frequency information truly facilitates statistical reasoning and decreases the rate of judgmental errors. It seems that the impact of frequency formats may interact with other factors, such as the experimental design (Kahneman & Tversky, 1996) and the transparency of the logical relation between the conjunct and the conjunction (Kahneman & Tversky, 1996; Mellers et al., 2001; Sloman, Over, Slovak, & Stibel, 2003; Wedell & Moro, 2008), making it difficult to draw unambiguous conclusions (also see McDowell & Jacobs, 2017 for a meta-analysis of the effect of frequency formats).

In sum, while challenging the results of the heuristics and biases approach, the frequentist approach also considers human probabilistic abilities as severely constrained to limited contexts. Frequentists argue that humans do not possess the ability to reason generally (e.g. about single-events), because it is impossible that evolution would have favored the selection of such general-purpose

reasoning mechanisms. Instead, the frequentist view states that humans possess an evolutionary ancient mechanism to reason about observed frequencies only. It is undisputed that humans, under certain conditions, *can* make use of observed frequencies to make rational judgments (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). The frequentist approach could, however, not convincingly demonstrate either that humans are reliably good intuitive statisticians as long as observed frequencies are available, nor that they are completely unable to draw inductive inferences when previous experience is not available. In parallel to the growing competition between the heuristics and biases advocates on the one side, and frequentists advocates on the other side, a third, yet again different view on human statistical abilities emerged: The Bayesian approach.

2.3. The Bayesian approach

Bayes´ theorem was named after Thomas Bayes, a Presbyterian minister and mathematician (1702-1761) and provides a method of determining conditional probabilities based on observations. One central difference between Bayesians and frequentists is their very definition of probability. While frequentists argue that probability is objectively defined over a specific reference class and refers to the relative frequency with which an event occurs, for Bayesians probability refers to a subjective degree of confidence. Conclusively, for Bayesians it is possible to assess the probability of a single-event by expressing one`s confidence that it will or will not occur. Referring to the example above, for frequentists there is no point in stating that "tomorrow it will snow 30%". For Bayesians, however, this statement corresponds to a certain degree of belief, namely, our quite confident intuition that there will not be any snow tomorrow. The interpretation of subjective believes as degrees of probabilities of future events has the inevitable consequence that the Bayesian understanding of probability is intrinsically subjective, which leads frequentists to the conclusion that the Bayesian interpretation of probability is "useless for scientific purposes" (Fisher, 1951, p. 7). The contrasting views on probability per se are connected to more far reaching and general philosophical questions regarding the nature of the human mind (see, e.g. Carnap, 1945; Gigerenzer, 1994; von Plato, 1994). While it is clearly beyond the scope of this dissertation to discuss and take a side on these issues, I will briefly introduce the basic idea of the Bayesian approach within this section and discuss its implications for the discussion on human statistical abilities.

As already mentioned, one central assumption of the Bayesian framework is that degrees of belief (e.g. that a certain hypothesis is true), can be expressed as probabilities, i.e. as real numbers between 0

and 1 (0 would mean the hypothesis is entirely false, 1 that it is completely true). To compute degrees of belief as probabilities, one needs two components: The prior probability and the likelihood. The prior probability reflects the degree of belief in a specific hypothesis prior to (or independent of) the newly observed data. The likelihood depicts the probability with which one would expect to observe the current data if the hypothesis in question was true. Combined, prior probability and likelihood yield the posterior probability, which is the updated prior probability conditional on the observed data. Bayes´ rule, i.e. the corresponding mathematical equation, ensures that the posterior probability of all hypothesis under consideration for explaining the observed data sum up to 1, with the most plausible hypothesis yielding the highest value (see, e.g. Perfors, Tenenbaum, Griffiths, & Xu, 2011 for a review and more detailed explanation of Bayes rule).

Consider the following practical example (taken from Tenenbaum et al., 2011, p. 1280): Imagine we observe John coughing (newly observed data), and we consider three competing hypotheses as explanation: h1= John has a cold; h2= John has lung cancer; h3= John has heartburn. The prior probability, i.e. our belief that John had any of these diseases before we observed him coughing, favors h1 and h3, because cold and heartburn are far more common diseases than lung cancer (i.e. in this case the prior probability equals the base-rate of the diseases). The likelihood, i.e. the probability that there is a causal link between the observed coughing and any of the explanations, favors h1 and h2, because only colds and lung cancer cause coughing, but not heartburn. Since h1 is the only hypothesis which scores high in both prior probability and likelihood, following Bayes´s rule, h1 is therefore weighted as the most plausible hypothesis. Hence, our intuition which told us from the beginning of this example that John´s coughing is probably caused by a cold, can be explained as Bayesian inference.

Over the last decades, Bayesian computational models have been used to explain various cognitive processes, ranging from aspects of basic, unconscious processing, such as perception (Yuille & Kersten, 2006), memory (Shiffrin & Steyvers, 1987; Steyvers, Griffiths, & Dennis, 2006), and sensorimotor systems (Körding & Wolpert, 2004), to aspects of higher level cognition such as diagnostic and conditional reasoning (Krynski & Tenenbaum, 2007) and predictions about the future of everyday events (Griffiths & Tenenbaum, 2006). More recently, Bayesian computational models have also been used to explain the tremendous learning efficiency of the developing mind of a child. For example, Xu and Tenenbaum (2007a) presented a Bayesian model of word learning, trying to explain how human learners can make meaningful generalization from limited examples of a new word. When confronted with a novel word with reference to an object, a child faces the problem that the word could refer to different levels of

categorization of this object. For instance, when it hears the new word "fep" in reference to a picture of a Labrador, the word could refer to a certain breed of dog, to all kinds of dogs, to animals in general, or even to all living things. Hence, the child has to decide which hypothesis about possible extensions of the label "fep" is best. If given only one example of a "fep", it should not have a strong preference for any of the hypotheses, though the more specific ones should be slightly favored. If given several examples of "fep", and they were all Labradors, it would be a very surprising coincidence if "fep" in fact referred to a more general category, such as all animals. When Xu and Tenenbaum (2007a) confronted human adults and children with a similar word learning task, they reacted just as predicted by the model: When they were given just one example of a "fep", they showed graded generalizations from lower level to higher level matches, i.e. from Labradors are "feps" to all living beings are "feps". When they were given three examples, by contrast, both adults and children made much sharper generalizations which were restricted to the lowest, most specialized level, i.e. the "Labradors are feps" level. Thus, meaningful generalizations from limited examples of a novel word´s referents possibly work through rational inferences which integrate prior knowledge about plausible word meanings with the statistical structure of the observed examples (Xu & Tenenbaum, 2007a; also see Xu & Tenenbaum, 2007b). In other studies, experimenters varied the evidence children saw about a causal system as well as the prior probability of several hypotheses about this particular structure, and they found that children typically chose the hypotheses with the greatest posterior probability in Bayesian terms (e.g. Bonawitz, van Schijndel, Friel, & Schulz, 2012; Bonawitz, Fischer, & Schulz, 2011; Gopnik, Sobel, Schulz, & Glymour, 2001; Sobel, Tenenbaum, & Gopnik, 2004).

One of the most crucial questions regarding the Bayesian approach to human inductive reasoning is how the prior probability is established and, even more far reaching, where our abstract background knowledge, allowing us to have priors and likelihoods, comes from in the first place. Is such abstract knowledge innate, for example in the form of innate concepts and core knowledge systems (e.g. Carey, 2009; Spelke, 1994)? Or is it acquired via associative learning mechanisms (e.g. Elman et al., 1996; Smith, 2001)? The Bayesian framework offers an "in between" explanation: Most likely there are some innate constraints, such as the whole object constraint (i.e. the fact that children learning words intuitively apply them to whole objects rather than parts (Markman, 1990)), and core systems of object representation, physics, psychology, and biology (e.g. Carey, 2009; Carey & Spelke, 1996; Spelke & Kinzler, 2007). In addition, there is probably also some higher-level abstract knowledge, learned by rational inferences itself

(e.g. Chater & Oaksford, 2008; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Tenenbaum et al., 2011).

One example of how such learning of inductive constraints (i.e. higher level abstract knowledge) could take place is the following scenario (taken from Goodman, 1955): Imagine there were several bags of colored marbles. We cannot see the content of the bags, but by drawing several samples from all of them we discover that each bag contains only one particular color (e.g. bag 1 only blue, bag 2 only red, etc.). If we now draw one marble from another bag in this population, and it is purple, it seems reasonable to conclude that subsequent draws from this bag will lead to purple marbles, too. The assumption that each bag is uniform in color is a learned inductive constraint, an "overhypothesis" that adds a piece to our abstract knowledge and enables subsequent generalizations. The learning of overhypotheses can also be modeled: The hierarchical Bayesian model (HBM) does not only choose among certain hypotheses, it also makes higher-order generalizations about these hypotheses. The top level of knowledge in an HBM is pre-determined (simulating abstract core-knowledge), and every level beneath can be learned and gets increasingly specific. Several HBMs have been computed to describe how category learning could work (e.g. Griffiths, Sanborn, Canini, & Navarro, 2008; Kemp, Perfors, & Tenenbaum, 2007; Navarro, 2006; Sanborn, Chater, & Heller, 2009). In one study (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002) children were presented with novel concepts and labels and rapidly learned the "overhypothesis" that not only chairs are organized by shape, but general categories of solid artefacts are, too. An HBM presented with the same data was able to make the same generalizations (Kemp et al., 2007; see Perfors et al., 2011 for a review of studies using HBMs).

Does this mean that the human mind is in fact a Bayesian statistical inference engine? How does this go together with all the empirically demonstrated limitations of human statistical reasoning previously described in this chapter? First of all, the Bayesian approach acknowledges that probability judgments, especially those requiring explicit manipulations of numerical values, can be influenced by heuristics and biased away from Bayesian norms (e.g. Tenenbaum et al., 2011). Second, even more importantly, the Bayesian approach is trying to give a rational framework of how human cognition in general and inductive inference in particular *could* work. The fact that in several cases the results of Bayesian models match those of actual human probabilistic computation gives reason to hypothesize that human statistical abilities might not be as limited and flawed as suggested by both heuristics and biases as well as frequentist approaches (e.g. Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Griffiths & Tenenbaum, 2006, 2009; Xu & Tenenbaum, 2007a, 2007b). The Bayesian approach provides a good

explanatory account of the ancient "inductive problem", i.e. the question how developing humans come to understand so much about the world from such sparse data by gradually and rationally revising their beliefs about the world in the light of new evidence.

But do these ideas truly reflect reality? Do young children, perhaps even before the onset of language, already possess and use statistical abilities? Are they able to draw statistical inferences without relying on general-purpose heuristics? Can they use statistical information to make rational predictions under uncertainty, also regarding single-events? Moreover, is their statistical inference mechanism domain general and can hence flexibly integrate various types of core knowledge, a capacity that would be necessary to learn overhypotheses and to accumulate higher-level abstract knowledge? In the following section I will review empirical developmental studies investigating these fundamental statistical abilities in human infants.

2.4. The ontogeny of intuitive statistical reasoning

Among the first psychologists to investigate the development of intuitive statistics were Jean Piaget and Bärbel Inhelder (1975). In their classic book "The origin of the idea of chance in children" they described a series of studies that have been conducted with children between 4 and 12 years of age. Their general conclusion was that probabilistic reasoning develops in three developmental stages. The first stage is the "intuition of rarity, but not of random mixture" (p. 97) at around age 4. Children of this age did recognize when an event was unlikely, but they could not explain why. For example, 4-year-olds expressed surprise when only items of the minority type were drawn from a population, but they did not seem to have a grasp of how specific proportions affect the outcome of random draws; at least they were not able to verbalize such an understanding. This only started around age 6, but even then, it remained inconsistent. Reaching a reliable understanding of the relation between a population´s proportion and randomly drawn samples is the characteristic of the second stage, called "chance and probabilities" (p. 103). This stage applies to children at around age 7. At this age, tested children could immediately explain why it was unlikely to randomly draw an item of the minority type of a population. They, however, did not yet realize that probability of an event can be expressed as a fraction of this event and the sum of the other events. At around age 12, children reach the last stage, "quantifications of probability" (p. 106). Tested children of this age had acquired an advanced understanding of probability. They could, for example, explain how multiple draws or varying ratios influence probabilities.

Thus, according to Piaget and Inhelder (1975), the development of statistical intuition is rather slow, and a true understanding of the relation between samples and populations does only emerge between seven and twelve years of age. However, as with many of Piaget´s studies, they relied heavily on verbal abilities, leaving uncertainty as to whether younger children´s failure reflects true limitations of probabilistic understanding, or rather simply constitutes a lack of ability to understand the task or the question, or a lack of ability to verbally express their thoughts. Indeed, over the last 60 years, more and more studies accumulated evidence that, when verbal demands are reduced, children reach Piaget and Inhelder´s stage 2 and 3 much earlier than originally assumed (e.g. Acredolo, O'Connor, Banks, & Horobin, 1989; Goldberg, 1966; Reyna & Brainerd, 1994; Yost, Siegel, & Andrews, 1962). In more recent years, new measures were established that entirely got rid of the tasks´ language component, allowing to test even younger children for their statistical abilities. In the following I will review these studies and their remarkable results. I will mainly focus on research on pre- or barely verbal infants, since those findings are most informative with regard to exploring a potentially innate inference mechanism which does not rely solely on experienced events and works without language or any form of formal education. I will only briefly mention work on older children where reasonable.

2.4.1. Reasoning from population to sample and assessing single-event probabilities

The majority of studies testing intuitive statistical inferences in pre-verbal infants use the violation of expectation (VOE) looking time paradigm. This paradigm uses the fact that infants look longer at scenes that violate their expectations, i.e. which they find surprising or unlikely. Accordingly, infants´ looking time, when compared between different statistically more or less likely scenes, can be used as dependent measure in order to assess infants´ intuitive understanding of statistical regularities. One of the first studies investigating intuitive statistics in infants used the VOE looking time paradigm to test 12-month-old infants for their ability to reason from population to sample (Teglas, Girotto, Gonzalez, & Bonatti, 2007). In this study, infants were first familiarized with a scene of four bouncing objects in a lottery machine: Three of them were yellow, one was blue. Subsequently, after a short occlusion phase, infants watched one of the objects exiting the lottery machine: In some trials it was one of the yellow objects, in others it was the blue object. Infants looked longer at the unlikely outcome of a minority object exiting the lottery machine, suggesting that at 12 months of age, infants already have an intuition about single-event probability. To exclude the possibility that infants simply reacted on the basis of a perceptual preference for attending to the singleton, the authors conducted a control experiment, in which a barrier

was placed in the lottery machine. This barrier separated both types of objects in a way that the single blue ball was below the barrier, close to the exit, whereas the three yellow balls were located above the barrier, and therefore were prevented from exiting the lottery machine. Now infants´ looking times showed the reverse pattern, indicating that they were not surprised to see the minority item exiting when it was the only possible outcome (also see Teglas, Ibanez-Lillo, Costa, & Bonatti, 2015 for similar results).

Within the same study, Teglas and colleagues (2007) tested pre-school children in a different task, requiring probabilistic inferences about another type of single-event. More specifically, the children´s task was to press a button when they saw a ball exiting a box. 3- and 5-year-olds acted more quickly when the ball exited from a side with three openings, compared to the opposite side with only one opening, suggesting that they had already intuitively shifted their attention towards the 3-hole-side. This indicates that they anticipated the more likely outcome. In accordance with Piaget´s and Inhelder´s (1975) findings, however, the 3-year-old children were unable to verbally express a judgment as to which side the ball was more likely to exit. In a later study (Teglas & Bonatti, 2016) the authors adjusted this paradigm for 12-month-old infants and recorded both looking time at either of the two outcomes as well as eye movements during a short occlusion phase prior to the exiting-event. The authors reported that infants looked longer at the unlikely event of the ball exiting on the 1-hole-side, indicating (posthoc) surprise. Interestingly, during the occlusion phase right before the ball exited, infants already moved their eyes towards the side with 3 holes, indicating that they also anticipated the ball to exit from there and hence that they constructed a forward representation of possible future events. Probabilistic intuition, therefore, seems to allow infants to prepare themselves for future events.

Xu and Garcia (2008) used the VOE looking time paradigm in combination with a live-demonstration to test 8-month-old infants in a task requiring inferences from population to sample. In this study, an experimenter showed the infants an open box containing a population of red and white Ping-Pong balls. For half of the infants, the population´s distribution of red to white balls was 70:5, for the other half it was 5:70. Subsequently, the box was covered and the experimenter drew, apparently randomly, a multi-item sample from it. In alternating trials, she removed either a 4:1 red to white or white to red sample. Similar as in the study by Teglas and colleagues (2007), infants looked longer at the scene in trials, in which the sample did not match the population´s distribution, i.e. when the mostly red sample was drawn from the mostly white population or vice versa. Control experiments ruled out that infants reacted to simple perceptual mismatches: When the experimenter removed the sample not from the box, but from her pocket, and thus the sample did not have a sampling relationship with the population

anymore, infants did not look longer at either of the outcomes. This indicates that infants in the original experiment were looking longer at the unlikely sample because it was improbable given the population´s distribution, and not just because of perceptual differences.

These studies (Teglas & Bonatti, 2016; Teglas et al., 2007, 2015; Xu & Garcia, 2008) demonstrated that at least by the age of eight months, infants can intuitively draw implicit inferences from populations to samples and have an intuition about single-event probabilities. The conclusions of these looking time studies have also been validated in studies applying an active-choice paradigm (modelled after a study by Feigenson, Carey, and Hauser, 2002). In one of them (Denison & Xu, 2010a) 12-to-14-month-old infants were presented with two transparent jars containing mixtures of pink (preferred) and black (non-preferred) lollipops. One consisted of 40 pink and 10 black ones, the other one of 10 pink and 40 black ones. The experimenter showed both populations to the infant before she randomly drew one lollipop out of each of the two jars (in a way that the infant could not see what was drawn) and placed the samples each in an opaque cup. Subsequently, the infant was encouraged to approach whichever cup she wanted. Infants chose the cup containing the sample drawn from the 40:10 (pink to black) population, indicating that they expected the random sample to be of the population´s majority type.

An alternative explanation for these results is that infants simply approached the sample drawn from the population with more desired objects, without necessarily considering their proportion. A similar absolute quantity heuristic could also explain the results of the previously described VOE looking time experiments. A later study of the same authors (Denison & Xu, 2014), therefore repeated the active-choice task with 10-to-12-month-old infants, but this time included a condition in which absolute and relative frequencies of preferred items were disentangled. More specifically, in this condition the population with the more favorable ratio of preferred to non-preferred items contained absolutely fewer preferred items (16:4 pink to black) than the other (24:96). Hence, if infants had used an absolute quantity heuristic, they should have approached the latter population, even though this population was less likely to lead to a pink lollipop as randomly drawn sample. Infants in this condition approached the sample drawn from the 16:4 population, indicating that they did use the proportions of lollipop types to guide their action. Similarly, another control condition ruled out the possibility that infants used an avoidance heuristic based on a comparison of absolute numbers of non-preferred lollipops. In this condition, both populations contained three types of objects: the previously used pink and black lollipops, plus neutral (i.e. intermediately preferred) green lollipops. One population consisted of a mixture of 8 pink, 8 black and 64 green lollipops, the other consisted of 8 pink, 12 black and 2 neutral lollipops. Infants intuitively preferred

the sample drawn from the latter population, indicating that they did not simply avoid non-preferred items, but instead compared the populations´ proportions in order to increase their chances of receiving a preferred item as randomly drawn sample[2].

Interestingly, however, a study with older children (Girotto, Fontanari, Gonzalez, Vallortigara, & Blaye, 2016) failed to find evidence for similar capacities in children aged between 3 and 4 years. These curious findings may be a first hint that the ability to reason about single-events develops in a non-linear fashion. They may, instead, also be explained by inhibitory control deficits in children aged between 3 and 4, which might have caused them to intuitively react towards absolute quantities of preferred items, instead of reasoning about proportions. For the infants in Denison and Xu´s studies (2010a; 2014), by contrast, these inhibitory control deficits may have had less severe consequences, since they were actively hindered from immediately approaching one of the jars by their parents. In any case, more research with different age classes is necessary to further investigate the developmental trajectory of the ability to reason about single-event probabilities involving inferences from population to sample.

The results of the described infant studies have two important implications: First, humans seem to possess the skill to use proportional information in order to reason from population to sample much earlier than previously assumed (Piaget and Inhelder, 1975), and this ability appears to be completely independent of the onset of verbal skills or any sort of mathematical education. Second, they revealed that the ability to reason about single-event probability was present in infants who had never experienced such events before. They therefore provide counterevidence to the frequentist position stating that humans can understand probabilities only as collections of experienced events (e.g. Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; see section 2.2. for discussion of the frequentist view). True statistical inferences, however, imply a genuine understanding of the relationship between populations and samples, i.e. a statistical inference mechanism should be able to draw inferences both from population to sample, as well as from sample to population.

---

[2] One could argue, however, that infants did not act optimally in this condition. Their chosen population was more likely to lead to a preferred type of lollipop than the other population. However, the likelihood for a non-preferred lollipop was even higher. Hence, the most likely outcome was a non-preferred lollipop for the population chosen by the infants, and a neutral lollipop for the neglected one.

2.4.2. Reasoning from sample to population

Infants´ ability to reason from sample to population has, to my knowledge, so far only been assessed using VOE looking time paradigms. In one study (Xu & Garcia, 2008), 8-month-old infants were presented with an opaque box, and watched an experimenter randomly drawing multi-item-samples from it. On alternating trials, she either removed 4 white and 1 red Ping-Pong balls from the box, or 1 white and 4 red Ping-Pong balls. Subsequently, she removed the box´s cover and revealed the contained population, which was a mostly white mixture for half of the infants, and a mostly red mixture for the other half. Infants reliably looked longer at the unexpected outcome, i.e. the population being mostly red and the sample being mostly white or vice versa. Therefore, when confronted with a multi-item sample, infants seemed to reason about the likely composition of the respective population. Again, control conditions ruled out that infants´ looking times were simply due to the perceptual mismatch between sample and population in the crucial condition. Another study used a slightly modified version of this task to test even younger infants (Denison, Reed, & Xu, 2013), and found that 6-month-olds, but not 4.5-month-olds succeeded. The authors suspected that the 4.5-month-olds´ failure was due to an inability to attend to the important parts of the scene, i.e. to the sampling event, which forestalled them from making accurate generalizations. Thus, it remains an open question, whether infants younger than 6 months would exhibit statistical reasoning abilities in an experiment with reduced task demands.

While the results of these two studies (Denison et al., 2013; Xu & Garcia, 2008) suggest that pre-verbal infants can generalize from sample observations, they, in contrast to studies investigating the reverse ability, did not include controls for absolute quantity heuristics. More specifically, it is possible, for instance, that infants simply attended to the absolute number of, e.g. red balls, and their looking times were guided by a "more is more likely" rule. Future studies, therefore, should test conditions in which absolute and relative frequencies are disentangled to examine whether infants truly use proportional information, rather than absolute quantity heuristics, in order to reason from sample to population. Another alternative explanation for infants´ behavior could be that they were following a general-purpose heuristic known from research on adult statistical reasoning capacities: representativeness. In all but one of the described studies investigating reasoning capacities from population to sample and vice versa, infants could have used a mental shortcut like "the more a sample and a population resemble, the higher the probability that they have a sampling-relationship". The sole exception is the study by Teglas and colleagues (2007): Here, infants looked longer at a sample that did resemble the population but was impossible due to a physical barrier separating it from the opening of the container. However, it is feasible

that infants in this particular condition considered the container as two separate sub-containers, each filled with a separate population. If viewed this way, infants could have applied the representativeness heuristic: "The sample should resemble the population located in the sub-container with an opening".

In sum, developmental research has shown that pre-verbal infants are capable of making probabilistic inferences from population to sample and vice versa, and they seem to have an intuition about the probabilities of never experienced single-events. However, alternative explanations, in particular the application of a representative heuristic, could account for these findings. In case infants did rely on such mental shortcut, they should, like adults, exhibit predictable biases and errors. In the following section I will review studies which addressed this concern, and which extended the investigation of intuitive statistical reasoning by exploring the question whether infants can integrate knowledge from different cognitive domains into their statistical inferences.

2.4.3. Integration of intuitive statistics with knowledge from other cognitive domains

As described above, the Bayesian approach suggests that statistical inferences are likely to be part of a central learning mechanism, allowing an individual to draw inductive inferences and thereby to rapidly acquire new knowledge. A prerequisite for such an inductive learning mechanism is to be domain-general. Hence, an individual has to be able to not only consider proportional information, but also to flexibly take into account various types of domain knowledge in order to draw correct inferences. One central question arising from the previously described findings is, therefore, whether pre-verbal infants are able to integrate domain knowledge into their statistical inferences in order to judge whether an event is truly random and therefore probabilistic, or whether it is non-random and the outcome will be determined or influenced by other factors.

2.4.3.1. Integration of physical information

Research on the development of physical knowledge and reasoning has produced ample evidence that very young infants already possess a basic understanding of physical properties and regularities of objects. In accordance with the core principle of persistence (i.e. objects continue to exist as they are, both in space and in time (Baillargeon, 2008)) they, for instance, expect objects to follow rules of cohesion, boundedness, continuity and solidity (Aguiar & Baillargeon, 1999; Baillargeon, Spelke, & Wasserman,

1985; Newcombe, Huttenlocher, & Learmonth, 1999; Spelke, Breinlinger, Macomber, & Jacobson, 1992). The following studies examined whether infants can combine and integrate naïve physics with intuitive statistics in order to predict and judge the outcome of events.

The above described study using the lottery machine paradigm (Teglas et al., 2007) is one example for a situation in which infants flexibly decided whether the outcome will be probabilistic and therefore based on the proportions of the two object types, or whether it will be determined by a physical constraint, namely a barrier separating the two object types. In this study, infants seemingly applied their knowledge about solidity in order to judge whether the event will be random or not. These findings, however, left open the question whether infants are truly able to *integrate* statistical and physical information, or whether they are only able to flexibly decide which source of knowledge will determine the outcome. In a slightly modified version of the same paradigm, Teglas and colleagues (2011) tested 12-month-old infants for their ability to integrate spatiotemporal information in statistical inference. Again, the lottery machine contained a population of three yellow and one blue object, all of them bouncing in random patterns. In contrast to the previous study, this time the authors varied both the spatial arrangement of the objects right before the occlusion phase (i.e. either the single blue object or one of the three yellow objects was close to the opening), as well as the duration of occlusion (between 0.04 and 2 seconds). Infants´ looking times were of graded nature: When the occlusion phase was short, infants seemed to judge the situation based on the spatial arrangement prior occlusion, i.e. they expected that object to exit, which was in closest proximity to the opening, regardless of whether it was of the minority or majority type. When occlusion lasted long, infants disregarded the spatial arrangement prior occlusion, and instead expected one of the majority objects to exit. When occlusion duration was intermediate, infants´ looking times reflected both the object proportions and their distance from the opening. Hence, infants integrated information about the ratio of objects, their physical arrangement and occlusion time to judge the outcome of an event. The authors computed a Bayesian model to assess the response of an ideal observer and found that infants´ looking time pattern was consistent with that of an ideal observer embodying abstract principles of object motion (also see Lawson & Rakison, 2013, for another example in which 12-month-old, but not 8-month-old infants seemed to integrate spatial information into their statistical inferences to predict the outcome of an event, suggesting that this ability develops at some point between 8 and 12 months).

Using a similar paradigm as in Xu and Garcia´ s (2008) study, Denison and Xu (2010b) tested 11-month-old infants for their capacity to integrate knowledge about a cohesion constraint into their

statistical inferences. More specifically, infants were first familiarized to the fact that one type of objects-green balls- were immobile (and therefore could not be sampled) due to a piece of Velcro attached to them. Subsequently, infants watched an experimenter drawing either four yellow and one red ball or one red and four yellow balls from an occluded box. When the experimenter revealed the content of the box- a population of balls with a 5:4:1 green to red to yellow ratio- infants looked longer at the 4:1 yellow to red sample compared to the 1:4 sample. This suggests that infants applied the cohesion constraint of green balls and computed probabilities over the remaining sets of objects. Importantly, when a different group of infants was presented with the exact same scene, but without the prior demonstration of immobility of green balls, they looked equally long at both samples, indicating that both violated their expectations since neither of the samples reflected the populations´ proportions. Using the same experimental setup, infants in a later study (Denison, Trikutam, & Xu, 2014) were either again familiarized with a deterministic physical constraint (all green balls are immobile), or with a probabilistic constraint (most, but not all green balls are immobile). The samples drawn from the population were either 4:1 green to red balls, or 1:4 green to red balls. When familiarized with the probabilistic constraint, infants looked longer at the sample that reflected the proportional composition of the population (4:1), but was unlikely because of the physical constraint. When the constraint was deterministic, i.e. one type of objects was completely immobile, infants looked equally long at both matching and non-matching samples, indicating that they found both samples equally unlikely, because both contained the immobile green balls. Therefore, infants were able to integrate a stochastic physical constraint rule into their probabilistic inferences, using the cohesion constraint of green balls to adjust the base-rate of balls available for sampling in their statistical computations.

Results of these three studies (Denison et al., 2014; Denison & Xu, 2010b; Teglas et al., 2011) clearly demonstrated that infants flexibly considered their knowledge about naïve physics to compute rational statistical inferences. At the same time they provided evidence that infants did not use a mental shortcut to make these inferences, such as the representativeness heuristic: In Denison and Xu´s study, for instance, infants did apparently not expect the sample to resemble the population it came from (and thus be representative of it). They instead, just like an ideal Bayesian observer, used the physical constraint to adjust the base-rate of balls in the population (i.e. the prior probability), and were thus able to fully integrate probabilistic inference with physical reasoning to compute a posterior probability.

The previously described study (Teglas & Bonatti, 2016) involving a ball bouncing in a box with one exit on one side, and three exits on the opposite side, also found infants´ looking time to be both

modulated by probabilistic (exit on the 1-hole side was possible, but unlikely) and deterministic components (exit on the 3-hole side was impossible due to a physical barrier). More specifically, infants looked longer at the scene when the ball exited at the 1-hole side compared to the 3-hole side when both events would have been possible, and they showed the reverse pattern when the ball exited the 3-hole side despite the fact that a physical barrier blocked those exits. The infants´ anticipatory looking behavior, however, indicated that they only anticipated the outcome when it was the most probable one, but not when it was the only possible one. The authors interpreted these curious findings as follows: Infants may anticipate events by programming their behavior, but they only do so when they can expect a gain in knowledge. By contrast, when they already know what the next step will be, e.g. when a ball can physically only exit via one way, they find no reason in programming their behavior ahead of time, because they already possess core knowledge of object solidity (e.g. Spelke et al., 1992). When interpreted this way, the results of this study suggest that in their first year of life, infants already know how to allocate their cognitive resources skillfully. More research is necessary to further explore this hypothesis.

Together, the discussed studies (Denison & Xu, 2010b; Denison et al. 2014; Lawson & Rakison, 2013; Teglas & Bonatti, 2016; Teglas et al., 2007, 2011) indicate that, at least by the age of 11 months, infants are capable of making statistical inferences that require the integration of naïve physics in order to compute a posterior probability.

2.4.3.2. Integration of psychological information

Results paralleling those just described for the physical domain were also found for the social domain: Xu and Denison (2009) presented 11-month-old infants with a task similar to the one used by Xu and Garcia (2008): Infants watched an experimenter drawing samples (in alternating trials either five white balls or five red balls) from an opaque population of Ping-Pong balls, and, as soon as the population´s content (mostly red balls) was revealed, the infants´ looking time at the scene was measured. In this study, infants were assigned to one of three conditions: In the random sampling condition, the experimenter drew samples randomly. Here, infants looked longer at the sample that did not match the population´s distribution (i.e. the white sample), replicating previous results (Xu & Garcia, 2008). In the other two conditions, infants first saw a short demonstration phase. In this demonstration, the experimenter explicitly expressed a preference for one type of balls (e.g. white), by picking only white balls from a 50/50 set. Subsequently, in the test phase, she drew the sample from the opaque box. The difference between

27

the two conditions was that in the test phase of the non-random sampling condition, the experimenter looked into the box while sampling, whereas in the blindfold condition she drew blindly. Infants in the non-random condition looked longer at the sample containing only red balls, presumably because this sample was incongruent with the experimenter´s goal. In the blindfolded condition, by contrast, infants looked longer at the only white sample, apparently realizing that, although the experimenter expressed a preference for white balls, she could not act according to her preference because of her lack of visual access. Hence, infants´ expectation of a random sampling event was overridden when a biased experimenter drew with eyes open, but not when she drew with eyes closed. These results suggest that infants flexibly considered intuitive psychological knowledge to judge the sampling conditions and drew according statistical inferences.

Infants´ sensitivity to sampling conditions was further demonstrated in a study with 15-month-old infants (Gweon, Tenenbaum, & Schulz, 2010). Here, infants were presented with a transparent box containing a large population of rubber balls in either a 4:1 (group 1) or 1:4 blue to yellow ratio (group 2). In both groups, infants witnessed the experimenter removing three blue balls from the population and demonstrating that all of them squeaked when squashed. Subsequently, the infant was handed over a yellow ball. Infants who saw the 4:1 blue to yellow population were more persistent in trying to make the yellow ball squeak, compared to those infants who saw the 1:4 population. Hence, infants in the latter condition seemingly inferred that the sample containing three (blue) minority objects was more likely to have been removed intentionally in order to demonstrate their special squeaking property, and accordingly generalized the squeaking property to blue balls only. In the 4:1 condition, by contrast, the three blue objects were of the majority type, and were therefore more likely to have been drawn randomly. Therefore, infants had less reason to assume that the yellow balls would not squeak, too and therefore generalized the squeaking ability to the population as a whole. Impressively, when the experimenter only drew a single squeaking blue ball out of either of the populations before handing over a yellow one to the infant, the infants showed no difference in persistence to making the yellow ball squeak. This suggests that they intuitively knew that, while drawing three blue balls in a row out of a mostly yellow population is quite unlikely, drawing a single blue ball is certainly possible and does not give any reason to assume a non-random drawing event. Conclusively, infants already seem to possess some intuitive appreciation of sample sizes.

A recent study further demonstrated that infants can use statistical information to infer an agent´s preferences (Wellman, Kushnir, Xu, & Brink, 2016). First, in a demonstration phase, 10-month-old infants

viewed an experimenter sampling five blue balls from a visible population consisting of either 80% blue balls (majority condition) or 20% blue balls (minority condition). Subsequently, in a new test scenery, the infants saw the same experimenter positioned between two transparent bowls: one containing only blue balls, the other containing only red balls. Infants in the minority condition looked longer at the scene when the experimenter picked a ball from the red bowl, seemingly violating their expectation that this person has a preference for blue balls and should act accordingly. Infants did not do so in the majority condition, indicating that they did not infer a preference from the initial demonstration event, since it was in accordance with the random sampling assumption. Importantly, the authors tested a third experiment, which was similar to the minority condition, but this time the experimenter sampled the five blue balls using a scoop. Thereby she unambiguously demonstrated that this unlikely event happened unintentionally. Accordingly, infants in this condition did not look differentially in test trials where the experimenter sampled blue or red balls. Hence, infants did not simply react to a difference in the experimenter´s behavior in the test compared to the demonstration phase; they instead apparently inferred a causal intentional state from a statistical pattern. (For similar findings with 20-month-old infants, toddlers and preschool children see Kushnir, Xu, & Wellman, 2010; Ma & Xu, 2011).

Together, these three studies demonstrated that already at 10-11 months, infants use information about others´ intentions in order to judge whether drawing is random or not. Vice versa, they are also able to use a violation of the random sampling assumption to draw conclusions about others´ intentions and preferences. Thus, infants can flexible integrate knowledge about naïve psychology with statistical information in order to draw meaningful conclusions about the social world and form correct predictions about others´ actions.

2.5. Conclusions

Despite a long history of research demonstrating the limitations and weaknesses of humans´ probabilistic abilities, more recent developmental work revealed that even pre-verbal infants possess basic intuitive statistical skills and have a sophisticated understanding of the relationship of samples and populations: When seeing a population and a sampling process, infants draw conclusions about the resulting sample. When watching a sampling process yielding a sample, infants reason about the population the sample was drawn from. Finally, when infants are confronted with a population and see the resulting sample, they draw inferences about the sampling process (see Figure 1 for a schematic illustration of the triad between

populations, sampling processes, and samples). Moreover, pre-verbal infants can reason about never experienced single-event probabilities in accordance with Bayes rule. This indicates that they possess an intuitive sense of probability which they apply in uncertain situations, and this sense works without any informative past experience. Infants do not only show implicit signs of surprise when confronted with unlikely events, they also anticipate probabilistic outcomes and use their intuitive sense of probability to guide their actions. Furthermore, infants can integrate substantive domain-knowledge into a domain-general probabilistic inference mechanism and can reason both about psychological and about physical variables when making statistical inferences. Relatedly, infants seem to have a random sampling assumption; if this random sampling assumption is violated, infants adapt their expectations accordingly.
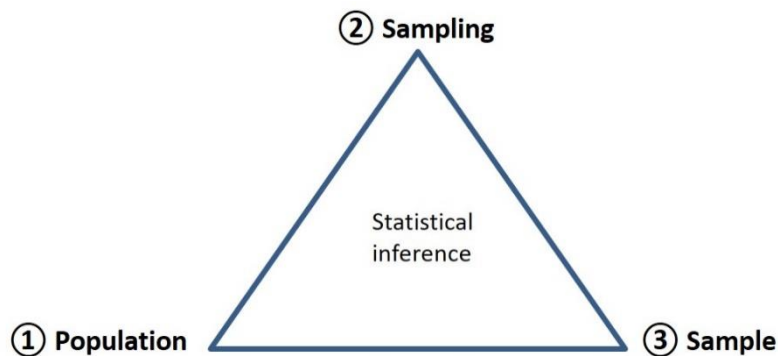
Figure 1 Schematic overview depicting the three possible directions of statistical inferences and evidence for their occurrence in human infants.

Apart from the discussed literature on infants´ intuitive statistical inferences, there is also a great deal of research suggesting that infants and young children are extremely sensitive to statistical regularities in their environment and make use of statistical input in the context of, for example, word and scene segmentation (Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996), language learning (Lany & Saffran, 2010; Xu & Tenenbaum, 2007a, 2007b) and causal reasoning (Kushnir & Gopnik,

2005, 2007; Schulz, Bonawitz, & Griffiths, 2007). Together, these findings suggest that infants are equipped with an efficient and powerful domain-general statistical inference mechanism from very early on that guides their learning and helps them acquire their rapidly growing knowledge about the world. Contrasting previous research on adults, infants´ inferences do not seem to be guided by simple heuristics, neither are they based on simple mechanisms detecting observed frequencies of events. This indicates that, in sharp contrast to hypotheses derived from both the frequentist and the heuristics and biases approaches, at least in their most basic and naïve form, human intuitive statistics are a sophisticated and well-functioning tool kit adapted to draw general conclusions from sparse data and to use these generalizations to predict the outcome of events. Importantly, this form of intuitive statistics seems to be completely unrelated to language- children younger than 4 years correctly anticipate the most likely outcomes, use these anticipations to guide their actions and are surprised when they do not come true- but they are unable to verbally express these intuitions. One might even speculate that accurate implicit naïve statistical inferences may be overridden by mental shortcuts later in development, perhaps with the onset of language and symbolic knowledge, as well as with formal mathematical education (see Cesana-Arlotti, Téglás, & Bonatti, 2012; Denison & Xu, 2012 for discussion of similar hypotheses). However, also for human adults, a slowly growing body of research suggests that statistical reasoning may not be as irrational as previously thought (e.g. Chater & Oaksford, 2008; Griffiths & Tenenbaum, 2006; Tenenbaum et al., 2011).

The finding that humans seem to have an innate statistical inference mechanism, which works independently of language and education raises a crucial question: Is such an inference mechanism uniquely human, perhaps accounting for our unprecedented higher-level cognitive abilities and flexibility? Or is it part of our evolutionary heritage and we may share it with other animals, most notably our closest living relatives, the nonhuman great apes? In the following section I will review research indicating that great apes possess the basic prerequisite for statistical reasoning abilities, namely a sense for quantitative and numerical information.

3. The evolutionary origins of quantitative abilities

> *"Our Western societies, ever since Euclid and Pythagoras, have placed mathematics*
>
> *at the pinnacle of human achievements. We view it as a supreme skill that either*
>
> *requires painful education or comes as an innate gift. In many a philosopher´s mind,*
>
> *the human ability for mathematics derives from our competence for language, so*
>
> *that it is unconceivable that an animal without language can count, much less*
>
> *calculate with numbers"*
>
> (Stanislas Dehaene, 1997, p.28)

Numerical information is omnipresent in the world and numbers are an integral part for even the simplest tasks in our daily life. We use numbers to quantify and measure objects and magnitudes, such as the amount of flour in a cake recipe or the distance from one city to the other. Numbers help us to rank and order things, like a grade in school or our place in a line. And finally, numbers allow us to understand and interact with the world in extremely sophisticated ways. For example, higher mathematics allows us to construct impressive architectural structures, to create and program artificial intelligence and to build statistical models to predict future events or to understand the basic processes that govern the inner workings of the universe.

As Dehaene´s (1997) above statement demonstrates, for a long time in human history numerical abilities have been viewed as uniquely human and closely linked to language and education (see e.g. Descartes & Lafleur, 1960; Kant, 1781; Ross, 1908). And in fact, the largest part of our sophisticated number skills relies on our exact symbolic number system, which is culturally learned and which enables us to represent numbers precisely via counting and number words (Dehaene, 1992). However, in the middle of the last century, early ethologists found first evidence for basic quantitative abilities in rats, birds, and monkeys. One of the earliest successful projects leading to, at that time, surprising results was led by a famous ethologist, Otto Koehler, and showed that a trained raven (*Corvus corax*) could differentiate containers depicting five points from containers depicting other numbers, even when shape, size and location of the points was varied (Koehler, 1943). Roughly a decade later, Hicks (1956) used a similar paradigm to test rhesus macaques (*Macaca mulatta*). He found that monkeys could discriminate cards depicting three objects from distractor cards that had up to six objects painted on them and, just as for the raven, performance held even when different types of stimuli and different spatial arrangements were used, demonstrating a concept of "Threeness" in monkeys. Using an extended operant conditioning

task, (Mechner, 1958) trained rats (*Rattus norvegicus*) to press one of two levers a fixed number of times before switching to the second one to receive a food reward. The animals learned to press the first lever, e.g. four times in a row before switching. This also worked with higher numbers up to 16. A follow-up study revealed that rats were really using numerical rather than temporal parameters to estimate the required amount of presses (Mechner & Guevrekian, 1962). However, the rats´ counting always remained rather imprecise; even after a considerable amount of training, rats trained on the number 4 still at times pressed the lever three, five or six times in some trials (less than four times was rare, due to a penalty that rats received for early switching). For rats trained on number 16, responses even ranged from twelve to twenty-four (Mechner, 1958). Later studies demonstrated that animals could also use numerical information presented in different modalities: Church (1984) trained rats to press the first of two levers when they heard an auditory sequence of two tones or a visual sequence of two light flashes, and the second one when they heard four tones or saw four light flashes, respectively. To get insights in whether these two pieces of knowledge were coded separately from each other in the rats´ brains or rather encoded as a more abstract numerical rule, the rats were next presented with mixtures of auditory and visual stimuli. Intriguingly, subjects intuitively generalized their knowledge across modalities and pressed, e.g. the second lever after hearing two tones *and* seeing two light flashes. These four studies were among the first to convincingly demonstrate a "concept of number" in nonhuman animals, which seems to be independent of modality and showed that animals might even be capable of simple arithmetic operations. In addition, they already gave some insights into the cognitive representation of quantity and its signatures.

Until today, the assessment of animal numerical and quantitative abilities has become extraordinarily broad and we now know that basic quantity assessments are relevant for basically all species of animals ranging from insects (e.g. honeybees (*Apis mellifera*; Dacke & Srinivasan, 2008; Pahl, Si, & Zhang, 2013)) and fish (e.g. mosquitofish (*Gambusia holbrooki*; Agrillo, Dadda, Serena, & Bisazza, 2008; Agrillo, Piffer, & Bisazza, 2011; Dadda, Piffer, Agrillo, & Bisazza, 2009); anglefish (*Pterophyllum scalare*; Gómez-Laplaza & Gerlai, 2011a, 2011b, 2013), guppies (*Poecilia reticulata*; Piffer, Agrillo, & Hyde, 2012)) over amphibians (salamanders (*Plethodon shermani*; Krusche, Uller, & Dicke, 2010; *Plethodon cinerus*; Uller, Jaeger, Guidry, & Martin, 2003), frogs (*Bombina orientalis*; Stancher, Rugani, Regolin, & Vallortigara, 2015) and birds (e.g. pigeons (*Columba livia*; Emmerton, 1998), robins (*Petroica longipes*; Garland, Low, & Burns, 2012), chicks (*Gallus gallus domesticus*; Rugani, Regolin, & Vallortigara, 2008), grey parrots (*Psittacus erithacus*; Pepperberg, 2006)) to various species of mammals (e.g. bears (*Ursus*

americanus; Vonk & Beran, 2012), dolphins (*Tursiops truncatus*; Jaakkola, Fellner, Erb, Rodriguez, & Guarino, 2005; Kilian, Yaman, Fersen, & Güntürkün, 2003), horses (*Equus caballus;* Uller & Lewis, 2009), elephants (*Elephas maximus*; Perdue, Talbot, Stone, & Beran, 2012).

This widespread emergence of quantitative abilities across the animal kingdom reflects the clear evolutionary advantages such capacities bring in diverse areas of an animal´s life. Numerous studies using field observations and experiments confirmed the adaptive value of quantitative competences in a variety of different contexts. The perhaps most obvious advantage of the ability to compare quantities concerns foraging decisions: Being able to discriminate and select the larger of two amounts of food will maximize food intake and thereby directly increase survival rates (e.g. Stephens & Krebs, 1986). Not surprisingly, there are studies showing that wild animals intuitively select the larger of two amounts of food (e.g. wild robins, *Petroica australis*; Hunt, Low, & Burns, 2008). But quantitative abilities can also be applied in social contexts, for example to assess chances of winning a fight with conspecifics. Playback-experiments demonstrated that a variety of social species appear to be sensitive to the number of calls of individuals: For instance, chimpanzees (Wilson, Britton, & Franks, 2002), lions (*Panthera leo*; McComb, Packer, & Pusey, 1994), hyenas (*Crocuta crocuta*; Benson-Amram, Heinen, Dryer, & Holekamp, 2011), wild dogs (*Canis lupus familiaris*; Bonanni, Natoli, Cafazzo, & Valsecchi, 2011), and howler monkeys (*Aloutta pigra*; Kitchen, 2004) seem to be able to assess the quantity of opponents in intergroup conflicts, and are more likely to engage in social confrontations when their own group outnumbers that of their opponents. In the context of successful mating, bank voles (*Myodes glareolus*) adjust their reproductive strategy to the level of sperm competition, i.e. the more other males are present the more sperm is produced by each of them (Lemaître et al., 2011). Lastly, numerical abilities are probably highly beneficial in the context of anti-predator strategies. For example, redshanks (*Tringa tetanus*) are less likely to be predated when they are part of a larger flock, suggesting that individuals which are able to select the larger social group have a clear fitness advantage (Cresswell, 1994). Relatedly, different species of fish were shown to be able to place themselves into the larger of two present shoals to achieve greater safety (Agrillo et al., 2008; Gómez-Laplaza & Gerlai, 2011b, 2011b).

While all these examples highlight the fitness benefits quantitative abilities may bring, they are unable to control for the use of non-numerical cues, that is, individuals may (and most likely do) use information that is confounded with number (such as area, duration, volume, intensity of smell, etc.), but not on number per se to make decisions (also see section 3.3.1.2. for further discussion of this topic). To get deeper insights into the extent and foundation of nonhuman numerical abilities, it is therefore

essential to consider laboratory experiments carefully controlling for various types of confounding information.

In the following, I will review some of the most important studies regarding quantitative cognition in nonhuman primates. First, I will discuss research focusing on the cardinal aspect of number, i.e. the ability of great apes and monkeys to estimate and judge absolute and relative quantities. Second, I will summarize studies investigating whether and how nonhuman primates can perform basic operations on their quantity representations, namely addition and subtraction, and whether they can represent quantities as proportions. Lastly, I will discuss what we know about the cognitive mechanisms and neurobiological foundation of numerical abilities. Considering the impressive breadth of the field, the aim of this chapter is not to be an exhaustive literature overview. Rather, I selected some of the most important studies that are worth to be mentioned with regard to the topic of this dissertation. Due to this criterion, I forwent to include research focusing on the ordinal aspect of number, that is "the ability to place objects in a series on the basis of some quantitative property (e.g. from short to tall) and, perhaps, to identify certain places in the series (e.g. the third place, including such things as placing group mates in dominance hierarchies)" (Tomasello & Call, 1997, p. 137) Also, I did not include literature demonstrating counting-like abilities in animals. For more exhaustive literature reviews see, e.g. Beran (2017) or Kadosh and Dowker (2015).[3]

3.1. Estimating and comparing quantities

The capacity of animals to estimate and judge quantities can be either assessed as absolute or as relative quantity judgments. Studies investigating *absolute* quantity judgments usually train the animal to choose stimuli containing a certain quantity of items over stimuli depicting any other quantity. The above-mentioned study by Hicks (1956) is an early example for such research. Importantly, the author controlled for the use of other non-numerical features like spatial configuration or surface area and could therefore convincingly conclude that the monkeys were able to judge the numerousness of objects. Until today, similar matching-to-sample tasks have been conducted with a variety of species, among them orangutans

---

[3]I will use the following nomenclature: The terms quantity and amount refer to any numerical *or* non-numerical magnitude. In many studies subjects could rely on either of the two, therefore I will use the terms quantity and amount as synonyms in such cases. The terms numerousness and numerosity, by contrast, only refer to discrete, i.e. countable quantities, and not to continuous magnitudes. Both will be used here as synonyms, neither of those is meant to imply an underlying concept of number.

(Vonk, 2014) and rhesus macaques (Jordan & Brannon, 2006). Their findings demonstrated that monkeys and apes used numerical information to match the sample, and they did so not only with visually presented stimuli, but also with auditory stimuli: when monkeys heard a certain number of vocalizations they preferentially and intuitively looked at a video depicting the number of conspecifics that matched the number of vocalizations (Jordan, Brannon, Logothetis, & Ghazanfar, 2005). Similarly, they also chose a matching number of arbitrary items on a computer screen after hearing a certain number of sounds (Jordan, MacLean, & Brannon, 2008). Hence, nonhuman primates seem to be able to extract and match numerical information from different modalities. Another famous study demonstrated a chimpanzee´s capacity to use Arabic numerals to correctly label absolute quantities of items (Matsuzawa, 1985). In this study, language-trained[4] chimpanzee Ai was taught to associate Arabic numerals with the according number of objects. She could correctly label the number of any type of objects up to numbers of six (similarly, African grey parrot Alex could verbally label the number of up to six objects (Pepperberg, 1987)). These and other studies demonstrated that nonhuman primates do have some understanding of absolute cardinality.

The perhaps most frequently assessed numerical capacity in nonhuman animals is *relative* quantity judgment, which usually involves choosing sets of objects based on a more-than or less-than rule. As mentioned above, a large variety of species has demonstrated the ability to spontaneously choose the larger of two quantities of food (e.g. Hunt et al., 2008; Krusche et al., 2010; Perdue et al., 2012). Chimpanzee Lana, for example, was allowed to choose which of two quantities of cereals she wanted to receive (Dooley & Gill, 1977). Lana chose the larger of the two quantities when each set consisted of up to ten pieces. Her performance declined, however, when the magnitude of difference between the two quantities was small and the set sizes were close to the upper limit of ten pieces. Call (2000) tested three orangutans for their ability to select the larger of two sets of food (between one and six items each) when both sets could only be viewed one after the other. More specifically, the first quantity was presented in a container, and only after a lid was placed on this container, the second container holding the second quantity was shown and likewise covered with a lid. Only then were apes allowed to choose between the two covered containers. Apes successfully chose the one holding the larger amount of food, suggesting that they did not just use a perceptual mechanism to make their decision, but rather used some sort of mental representation. Adding on this finding, another study (Beran, 2001) investigated two chimpanzees´

---

[4] The term "language-trained" refers to individuals, who were trained to communicate with humans via sign language, lexigrams or labelled plastic chips.

ability to select the larger of two quantities when each item was presented separately, one at a time. More precisely, each food item was shown to the ape and then placed into one of two opaque containers. Only after both containers were fully baited (between one and nine items each), the chimpanzee was allowed to choose between them. Hence, subjects had to mentally keep track of the total quantity in each set. Again, they successfully chose the larger of the two quantities, suggesting that apes mentally represented each quantity, even though they were never entirely visible at once. Just as in previous studies, chimpanzees´ performance was strongly ratio dependent, with decreasing performance the smaller the relative difference between quantities was. Similar results were found for other great ape species: Hanus and Call (2007) demonstrated that bonobos, gorillas, orangutans, and chimpanzees chose the larger of two amounts of food, regardless of whether the two sets were presented simultaneously or sequentially (quantities 1 to 10). Again, the ratio between the two quantities reliably predicted the apes´ performance.

Apes are not only able to use visual, but also auditory information to judge quantities: Beran (2012) presented chimpanzees with an occluded version of the item-by-item task (Beran, 2001). Here, chimpanzees only heard individual items falling into different containers instead of seeing them and hence were prevented from visually tracking the amount of food. Again, subjects were able to pick the container holding the larger amount of food, provided that the ratio between quantities was large enough. Importantly, chimpanzees even succeeded with comparisons between a fully visible set and a sequentially presented auditory set, demonstrating that they did not use the duration of presentation as cue for the amount of food. These studies involving sequential presentation of food items (Beran, 2001, 2012; Call, 2000; Hanus & Call, 2007) demonstrated that great apes were truly using an enumerative process, rather than a simple perceptual mechanism that provides parallel apprehension of quantity ("subitizing", see Mandler & Shebo, 1982).

Similar studies have been conducted with different monkey species and lemurs (capuchin monkeys (*Cebus apella*; Addessi, Crescimbene, & Visalberghi, 2008; Beran, Evans, Leighty, Harris, & Rice, 2008; Evans, Beran, Harris, & Rice, 2009; vanMarle, Aw, McCrink, & Santos, 2006), baboons (*Papio anubis*; Barnard et al., 2013), lemurs (*Eulemur mongoz*: Lewis, Jaffe, & Brannon, 2005)) and found that they, like apes, chose the larger of two food quantities, regardless of whether they were presented simultaneously or sequentially, and all found a ratio dependent performance. These and other findings outside the primate order (e.g. Abramson, Hernández-Lloreda, Call, & Colmenares, 2013; Dadda et al., 2009; Garland et al., 2012; Hunt et al., 2008; Perdue et al., 2012) indicate that the capacity to form and compare

representations of food quantities developed very early in animal evolution, long before the emergence of human and nonhuman great apes.

Choosing the larger of two quantities of food seems to be an intuitive capacity, but are nonhuman primates also able to indicate the smaller of two quantities? Boysen and Berntson (1995) addressed this question with two laboratory-trained chimpanzees, Sheba and Sarah. In the test, both chimpanzees were allowed to choose between two quantities of candy. The crucial difference to previous studies was that the selected quantity was given to the other individual, and they were left with the one they had not indicated. Interestingly, both subjects had massive difficulties to learn to select the smaller of two food quantities in order to receive the larger one. However, Sheba had previously learned to associate quantities with Arabic numerals. When she was given the same task with Arabic numerals instead of food items, she suddenly did select the smaller Arabic numeral to receive the larger food reward (also see Boysen, Berntson, & Hannah, 1996). This study demonstrated one of the disadvantages of using food items as stimuli: Food is a very salient stimulus and, while it is clearly advantageous to induce intuitive responses that do not require any training, it may at the same time mask some of apes´ and monkeys´ competences because nonhuman primates have difficulties to inhibit their natural tendency to reach towards the larger food amount. Another study testing Old world monkeys (olive baboons and long-tailed macaques, *Macaca fascicularis*) in a quantity comparison task confirmed this presumption and found that the monkeys´ performance was enhanced in conditions in which inedible objects, instead of food items, were used (Schmitt & Fischer, 2011). Another disadvantage of using food items as stimuli is that the animals´ judgments in these studies were most likely not purely numerical. Instead, individuals probably used a combination of different kinds of information, among them number, surface area, duration of stimulus presentation (in the case of sequentially presented food items) etc. to judge the relative quantity of food items. However, there are also studies demonstrating animals´ abilities to discriminate quantities of arbitrary objects, while controlling for other stimulus dimensions. For example, Thomas and Chase (1980) found that squirrel monkeys (*Saimiri sciureus*) could be trained to discriminate cards displaying arrays of two to nine dots. The monkeys learned to always choose the card depicting the smaller of two numerosities, while he controlled for spatial arrangement, area and brightness. A more recent study (Beran, 2007) tested rhesus macaques with a computerized version of the item-by-item task (Beran, 2001). Monkeys saw animations of items being sequentially dropped into two opaque containers and then reliably chose the digital container holding the larger quantity. Importantly, in contrast to previous non-computerized studies, here the author controlled for non-numerical cues such as duration and surface

area. Even more convincingly, another touchscreen study (Cantlon & Brannon, 2007a) showed that monkeys did in fact pay more attention to numerical information than to other properties of items. More specifically, monkeys were tested in a matching-to-sample task with more than one correct answer, i.e. the sample could either be matched in terms of its numerosity or in terms of a non-numerical property such as color, surface area or shape (e.g. two stars could be matched either with two squares or with four stars). Monkeys based their choice intuitively and consistently on numerical values, suggesting that they spontaneously and routinely represent numerical properties of their environment (but see Clearfield & Mix, 1999 for a study finding contrasting results with human infants).

In sum, studies on absolute and relative quantity judgments found evidence that nonhuman primates cannot only represent and discriminate quantities (i.e. magnitudes), they can also represent purely numerical information, when non-numerical cues are controlled for and they can extract numerical information from different modalities. Furthermore, there are also studies showing that monkeys and apes are capable of size judgments (Schmitt, Kröger, Zinner, Call, & Fischer, 2013; Menzel, 1961, Menzel & Davenport, 1962), judgments of other continuous quantities such as liquids (Beran, 2010; vanMarle et al., 2006), and they have some understanding of conservation of quantity (Call & Rochat, 1996; Muncer, 1983; Suda & Call, 2004, 2005). For the purpose of this dissertation these studies are less relevant than other sections of numerical cognition, therefore I will not go into further detail (please refer to, e.g. Beran, 2017 for a review).

3.2. Arithmetic operations

According to Gallistel and Gelman (1992, 2000) a true concept of number involves the ability to perform operations or procedures on numbers. In the following I will review studies investigating whether great apes and monkeys are capable to sum and subtract quantities and whether they can represent quantities as proportions or fractions.

3.2.1. Summation and subtraction

One of the earliest and most impressive examples for arithmetic abilities in a nonhuman animal is that of home-raised chimpanzee Sheba (Boysen & Berntson, 1989). Sheba not only learned to use Arabic numerals to label sets of any kinds of objects. She was also able to label summed sets of objects which

were distributed across different locations. More specifically, Sheba´s task was to move among three "foraging" sites, in two of which a small number of food items was hidden. After she had observed the food at each location, she was asked to return to a home base and pick the card with the Arabic numeral that was equal to the sum of the food items she had seen. Sheba was able to give the correct numeral label for the total number of items (for sums of up to four items), even though she had never seen them all at once. Moreover, she also passed the task when, instead of food items, she found Arabic numerals at each side.

Since this early research, a variety of different methodologies have been applied to study nonhuman primates´ ability to sum and subtract quantities. One type of studies used the violation of expectation looking time paradigm, which was originally developed to investigate human infants´ capacity for basic arithmetic operations (Wynn, 1992). In a study by Hauser, MacNeilage, and Ware (1996), for example, free ranging rhesus macaques saw certain quantities of food items, e.g. two eggplants, on a stage. Then, the experimenter blocked the subjects view with a screen and visibly added or removed, e.g. one eggplant to or from the stage. Subsequently the screen was removed and the subject saw the outcome. Results showed that monkeys looked longer at impossible outcomes, e.g. 2-1=2, compared to possible outcomes, e.g. 2-1=1 (also see Hauser & Carey, 2003). In a similar study (Flombaum, Junge, & Hauser, 2005) rhesus macaques also looked longer at impossible addition and subtraction outcomes when they contained larger quantities (e.g. 4+4=4 vs. 4+4=8). Thus, monkeys seem to be able to track the total number of objects behind a screen when subsets of items are added or removed.

Another type of studies used more explicit choice tasks to investigate great apes´ ability to perform basic arithmetic operations. For instance, two language-trained chimpanzees were presented with two sets of chocolate and were allowed to choose which of the sets they wanted to consume (Rumbaugh, Savage-Rumbaugh, & Pate, 1988; Rumbaugh, Savage-Rumbaugh, & Hegel, 1987). Subjects chose the larger quantity of chocolates, even when each of the two sets was split in a pair of food wells. For instance, they preferred a 4+3 pair over a 5+1 pair, indicating that they compared the total quantity instead of choosing the set containing the largest subset. Importantly, however, in this study numerical information was confounded with spatial volume (as often in studies using food items as stimuli), hence it remains an open question whether subjects truly added individual numbers, or whether they rather compared the total volume of food items in both sets. More recent studies controlled for such perceptual cues by presenting items sequentially, without the subject ever seeing the two sets to be compared in their entity. For example, in the previously described study (Beran, 2001), chimpanzees watched while

one to six food items were placed in each of two opaque containers. Subsequently, another one to six food items were added to each of the two containers. Chimpanzees reliably chose the cup holding the larger quantity of food, even when each quantity was presented in three subsets (e.g. 2+2+3 vs. 3+4+1), or when one item was removed from one of the sets before the chimpanzee made her choice. A similar study found comparable results for orangutans (Anderson, Stoinski, Bloomsmith, & Maple, 2007; also see Anderson et al. (2005) for a study on gorillas).

A different approach to investigate nonhuman primate arithmetic without using food items as stimuli is applying touchscreen setups. For example, Cantlon and Brannon (2007b) tested rhesus macaques and humans in the same matching-to-sample task: Subjects were shown a sample image and, after a short delay, a second sample image. Subsequently, they were offered a choice between two arrays, one of which was the sum of the two sample arrays. Both humans and monkeys succeeded in this task, even though non-numerical cues like cumulative surface are were controlled for. And, interestingly, performance of monkeys and humans followed very similar patterns: Both showed a ratio-dependent performance, i.e. increasing accuracy with larger, compared to smaller relative differences between quantities.

3.2.2. Understanding proportions

Another interesting arithmetic operation that has attracted less attention in comparative research compared to summation and subtraction is understanding proportions. Humans frequently make use of proportional information in various everyday situations, for example to calculate the discount price of a dress during summer sales or when we mix a cocktail using 1/3 alcohol and 2/3 juice. Despite the fact that developmental research has shown that children have difficulties to learn the meaning of proportional information (e.g. Hartnett & Gelman, 1998; Siegler, Fazio, Bailey, & Zhou, 2013) comparative research has produced some evidence that nonhuman primates have a basic understanding of proportions.

One of the most famous examples in this field stems from an early study with language-trained chimpanzee Sarah demonstrating an understanding of abstract number concepts (Woodruff & Premack, 1981). In a matching-to-sample task, Sarah was first trained, among two available objects, to select the one that was physically identical to a third one. Once the ape had mastered this simple task, she was tested in more abstract versions, with the two choice stimuli being physically different from the sample stimulus, but one of them depicting the same fraction. For instance, the sample was a glass half-filled with

liquid, and the subject could choose between half an apple and three-quarters of an apple. The chimpanzee matched the half apple to the half-filled glass and did the same with other pairings of the fractions 1/2, 3/4, and 1/4, independent of the type of the stimuli to be paired. The authors also tested four younger non-language-trained chimpanzees, all of which failed in the abstract version of the task, suggesting that language training may have facilitated Sarah´s learning of magnitude relations.

A general understanding of proportions, though, does not depend on language. Vallentin and Nieder (2008) tested rhesus monkeys and human adults in spatial proportion discrimination task. In a delayed matching-to-sample paradigm, proportions were presented as two horizontal bars. The relation of the length of the two bars specified one of four proportions: 1:4, 2:4, 3:4 or 4:4. Monkeys learned to discriminate these proportions at a precision comparable to that shown by human participants. Moreover, monkeys could transfer the learned discrimination rule to new stimuli with novel ratios (3:8 and 5:8), demonstrating an abstract understanding of proportionality. Another study tested rhesus macaques in a task in which proportions were specified as numbers of items (Drucker, Rossa, & Brannon, 2016). Monkeys were trained in a touchscreen setup to choose one of two arrays, namely the one with a greater ratio of positive to negative stimuli, in order to receive a reward. After the training, subjects were able to transfer this proportional rule to new stimuli, regardless of the absolute number of stimuli in each array. Importantly, the authors controlled for the use of non-numerical cues, such as surface area. Moreover, both accuracy and reaction time were modulated by the ratio between ratios: The higher the ratio between the two proportions, the faster and better was the monkeys´ response. Very similar results have been found in a study testing human infants in a looking time paradigm (McCrink & Wynn, 2007). After habituating them with multiple examples of a single ratio, infants were able to discriminate between new examples of this ratio and novel ratios. Again, performance was ratio-dependent; infants responded more accurately, the higher the ratio between ratios was, regardless of the absolute number of stimuli within the arrays.

The striking similarity of human and nonhuman performance patterns in this and other non-verbal quantity-related tasks, clearly suggest not only that quantitative abilities are an evolutionary ancient capacity, but also that they seem to share the same cognitive foundation in humans and nonhuman primates.

3.3. The cognitive foundation of quantitative abilities

As the previous parts of this chapter demonstrated, nonhuman primates share with humans the capacity to represent numerical information from different modalities and are even able to perform mathematical operations on these representations. This raises a fundamental question: What is the cognitive foundation of these capacities? Two candidate mechanisms have been proposed and are frequently assessed in comparative studies: The analogue magnitude system (or approximate number system, see section 3.3.1.2.) and the object file system. In the following I will explain the two and discuss the evidence we have to date for each of them.

3.3.1. Analogue magnitude system

3.3.1.1. Signatures and empirical evidence

Many of the previously described studies investigating quantitative abilities in a variety of different species and contexts revealed striking and stable parallels in the performance of human and nonhuman subjects: Performance in quantitative tasks is generally not exact, but approximate, with increasing inaccuracy the larger the quantity to be represented. Recall the initially mentioned experiment on rats (Mechner, 1958). The rats learned to press a lever a fixed number of times, but their quantity assessment was imprecise: When trained on number four, responses ranged from three to six; when trained on 16, the rats pressed the lever between twelve and 24 times (note that subjects tended to over- rather than underestimate the number of presses because underestimations were punished). The apparently increasingly imprecise representation leads to characteristic signatures in quantity discrimination tasks: Performance generally improves, the larger the ratio between the two quantities to be discriminated. For example, when humans and rhesus monkeys were tested in the same matching-to-sample touchscreen task, both their accuracy and their reaction time were likewise modulated by the ratio of the two quantities to be compared (Cantlon & Brannon, 2007b), with increasing performance and faster responses the larger the relative difference was. Similar ratio dependent performances were found in a wide range of studies requiring subjects to compare quantities (e.g. Beran 2001, 2012; Dooley and Gill, 1977; Hanus & Call, 2007; Perdue et al., 2012) or proportions (e.g. Drucker et al. 2016; McCrink & Wynn, 2007). These conserved signatures of quantity representations have been traced back to a common cognitive mechanism- the analogue magnitude system.

The analogue magnitude system represents number (and also other magnitudes like duration or space) by a mental magnitude that is roughly proportional to and thus a direct *analogue* of the number of items in the set being enumerated (see, e.g. Carey, 2009; Dehaene, 1997). It thus enables individuals to estimate and compare arbitrarily large quantities, but only in an approximate way (e.g. Nieder & Dehaene, 2009). Its accuracy follows Weber´s Law: Discriminability of two sets varies as a function of the ratio of the set sizes to be compared, independently of their absolute numerosity (e.g. Cantlon & Brannon, 2006, 2007b). For example, if a subject can discriminate 2 from 4 objects, it is also able to discriminate 10 from 20 or 500 from 1000. This fundamental characteristic yields specific signatures that can be used to identify the involvement of this system in cognitive tasks. The fact that these signatures have been found in a large variety of species, ranging from fish (Buckingham, Wong, & Rosenthal, 2007) and birds (Ain, Giret, Grand, Kreutzer, & Bovet, 2009; Rugani, Cavazzana, Vallortigara, & Regolin, 2013) to monkeys (Barnard et al., 2013; Cantlon & Brannon, 2007b), and great apes (Beran, 2004; Call, 2000; Hanus & Call, 2007), clearly suggests that the analogue magnitude system is an evolutionary ancient mechanism (see Beran, 2017 for a review).

In humans, the analogue magnitude system is present from birth (Izard et al. 2009) and its accuracy seems to improve over development (Halberda & Feigenson, 2008; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Libertus & Brannon, 2010; Odic, Libertus, Feigenson, & Halberda, 2013; Xu & Spelke, 2000). One open question that is still debated for both human and nonhuman species is whether the scaling that occurs in quantitative representation through the analogue magnitude system is linear, with increasing variability for larger magnitudes (e.g. Brannon, Wusthoff, Gallistel, & Gibbon, 2001; Gibbon & Church, 1981; Whalen, Gallistel, & Gelman, 1999), or whether it is logarithmic, with a logarithmic compression of the mental number line which is used for numerical representations (e.g. Dehaene, 2003; Roberts, 2006; also see Cantlon, Cordes, Libertus, & Brannon, 2009 and Dehaene, Izard, Spelke, & Pica, 2008 for discussion).

3.3.1.2. Analogue magnitude vs. approximate number system

While there is general consensus about the existence of the analogue magnitude system and its involvement in quantitative judgments (see, e.g. Carey, 2009 for a review), there are still some highly debated issues regarding its exact nature. One central question is whether this system is restricted to numerosity alone and can therefore be described as an approximate number system (ANS) in the more

narrow sense. Alternatively, it could be a system with broader reach, representing all kinds of discrete and continuous magnitudes. This is related to the broader debate of whether numerical cognition is a quintessential cognitive domain with a specialized cognitive mechanism or whether the analogue magnitude system is part of a broad domain in which all quantitative dimensions share computational mechanisms (see, e.g. Cantlon, Platt, & Brannon, 2009; Lourenco, 2015 for reviews).

In virtually every naturally occurring situation numerical information is confounded with at least one continuous factor. For example, the longest line at the supermarket counter is usually also the one in which most people stand. Similarly, the bowl containing the larger number of cherries, is most likely the fuller one. Relatedly, the group of chimpanzees numbering the most members will also be the loudest one. So, does it, from an evolutionary perspective, even make sense to develop distinct cognitive systems for numerical and non-numerical magnitudes? Also, numerical and non-numerical magnitudes such as duration and surface area share some structural similarities, namely their analogue format. This leads to approximate representations with signatures obeying Weber´s law, as cognitive research on humans has demonstrated. For instance, human adults exhibit very similar, ratio-dependent signatures when discriminating durations, distances or physical sizes, all following Weber´s law (e.g. Droit-Volet, Clément, & Fayol, 2008; Fias, Lammertyn, Reynvoet, Dupont, & Orban, 2003; Henmon, 1906). On the individual level, performance in numerical discrimination tasks seems to be correlated with performance in tasks requiring a discrimination of, e.g. surface area or line length, even when controlling for other factors, such as verbal competence (DeWind & Brannon, 2012; Lambrechts, Walsh, & Wassenhove, 2013; Lourenco & Bonny, 2014; Lourenco, Bonny, Fernandez, & Rao, 2012; but see Gilmore, Attridge, & Inglis, 2011). Similar as human adults, pre-verbal infants have been found to have the same discrimination thresholds for numerical and non-numerical magnitudes such as length of a line, spatial extent, and duration of a tone (Brannon, Lutz, & Cordes, 2006; Brannon, Suanda, & Libertus, 2007; de Hevia & Spelke, 2010; vanMarle & Wynn, 2006). The developmental trajectories of such continuous and discrete magnitude discrimination abilities seem to run parallel, with similar and simultaneous improvements of accuracy for, e.g. numerical values and duration (Bonny & Lourenco, 2013; Droit-Volet et al., 2008; Holloway & Ansari, 2008). Furthermore, some studies showed that pre-verbal infants generalize regularities across magnitudes (e.g. de Hevia, Izard, Coubart, Spelke, & Streri, 2014; de Hevia & Spelke, 2010; Izard, Sann, Spelke, & Streri, 2009; Srinivasan & Carey, 2010). For example, infants intuitively transferred an ordering rule from numerical values to line lengths: Infants who were familiarized with ascending numerosities, looked longer at descending compared to ascending line lengths and vice versa (de Hevia & Spelke, 2010).

Similarly, infants associated larger temporal durations to larger spatial extent (Srinivasan & Carey, 2010). All these factors speak in favor of an at least "partially overlapping" system for the processing of discrete and continuous quantities.

On the other hand, there are some studies finding contrary results, both regarding the developmental trajectory of numerical and non-numerical abilities, and regarding generalizations across magnitudes. Odic and colleagues (2013), for example, found that children´s acuity to discriminate surface areas improved with age, but was consistently higher than their number acuity, suggesting a potential difference in the underlying mechanisms that encode and/or represent approximate area and approximate number. DeWind and Brannon (2012) found that college students´ ability to judge relative numerosities of items clearly improved after having received some training trials. However, this numerosity training did not influence students´ ability to discriminate non-numerical magnitudes, such as line lengths, suggesting that numerosity and line length are represented in distinct systems. Hence, while it is undisputed that continuous and distinct quantities share some properties, most notably their analogue format, it remains still unclear whether and to which degree representations of both types of magnitude overlap.

In sum, nonhuman primates share with humans a sense of number and can extract discrete numerical information independently from continuous quantities (e.g., Beran, 2007; Cantlon & Brannon, 2007a). It remains uncertain, however, whether there is a distinct cognitive mechanism supporting only numerical judgments (i.e. an approximate number system), or whether the cognitive foundation of numerical judgments is the same, or at least overlapping with that for quantitative judgments involving other types of magnitudes (i.e. an analogue magnitude system). Due to this ambiguity I will use the more general term analogue magnitude system throughout this dissertation.

3.3.1.3. Relation of "primitive" system and formal mathematics

Another profound question that is still under debate is how this "primitive" evolutionary ancient system is related to human formal mathematical abilities (also section 3.4. for a neurobiological approach to this question). One hypothesis is that individual differences in higher mathematics are, at least in part, grounded in individual differences in the accuracy of the analogue magnitude system. A variety of studies found support for this hypothesis (Bugden & Ansari, 2011; Halberda, Mazzocco, & Feigenson, 2008; Holloway & Ansari, 2009; Libertus, Odic, & Halberda, 2012; but see Libertus, Feigenson, & Halberda,

2013b; Smedt, Noël, Gilmore, & Ansari, 2013). For example, Halberda and colleagues (2008) assessed the individual performance of adolescents in approximate numerical discrimination tasks and found a surprisingly large variation in accuracy of the analogue magnitude system. Interestingly, individual accuracy was correlated with symbolic mathematical performance repeatedly measured in early childhood: Children who scored well in a "test of early mathematical ability" (TEMA) while still in kindergarten were found to have a higher accuracy of the analogue magnitude system as a teenager, even when other factors (e.g. working memory) were controlled for. Similarly, vanMarle, Chu, Li, and Geary (2014) found that the accuracy of the analogue magnitude system was correlated with performance in tasks measuring symbolic quantitative knowledge in preschoolers. These and other studies suggest that early (and probably inherent) accuracy of the analogue magnitude system is predictive for later mathematical achievement, which supports the hypothesis that human formal mathematics are based on an evolutionary ancient system.

3.3.2. Object file system

Some studies investigating quantitative abilities in nonverbal tasks reported performance patterns and limitations that do not fit the signatures of the analogue magnitude system. More specifically, in some cases subjects were able to discriminate small quantities in an exact and ratio-independent manner. For instance, Hauser, Carey, and Hauser (2000) tested rhesus macaques in a quantity discrimination task with sequentially presented food items. The monkeys could easily discriminate two small quantities from each other (e.g. two versus three), but they failed to reliably discriminate small sets from large sets (e.g. three versus eight), even though the latter should be easier to be discriminated with regard to Weber´s law. As explanation for these finding it was proposed that humans and nonhuman species possess a second system for numerical representation: The object file system supports exact representation and discrimination of small sets (usually up to four items). This system represents objects in separate "files", which hold information about their identity and features, and the number of files can be accessed to generate quantitative information about the array as a whole (e.g. Feigenson et al., 2002; Uller, Carey, Huntley-Fenner, & Klatt, 1999). The system´s limitation to small set sizes is probably due to attentional and working memory limits, since only a small number of "files" can simultaneously be kept in short-term or working memory. The ability to precisely and rapidly represent small numbers of items has been revealed in some studies with human adults (e.g. Choo & Franconeri, 2014; Dehaene & Cohen, 1994; Kaufman, Lord, Reese, & Volkmann, 1949; Mandler & Shebo, 1982; Trick & Pylyshyn, 1994), human

children (e.g. Feigenson & Carey, 2005; Xu, 2003) and also some nonhuman animals, including fish (Piffer et. al, 2012), birds (Garland et al., 2012), monkeys (Barner, Wood, Hauser, & Carey, 2008; Hauser et al., 2000), and apes (Murofushi, 1997; Tomonaga & Matsuzawa, 2002). However, there are several studies in which less or no support for an object file system could be found, both in humans (adults: Balakrishnan and Ashby, 1991, 1992; children: Cantlon, Safford, and Brannon, 2010; Cordes and Brannon, 2009), and in nonhuman species (e.g. great apes: Hanus and Call, 2007; monkeys: Barnard et al., 2013; Beran, 2007; Cantlon and Brannon, 2006). Importantly, the presentation format (sequential vs. simultaneous) does not seem to influence which type of signatures are found: Hauser et al. (2000) and Beran (2007) both tested rhesus macaques with sequentially presented quantities, and, while Hauser and colleagues found limitations in accordance with the object file system (i.e. absolute set size restricted performance), Beran found typical analogue magnitude system limitations (i.e. relative difference restricted performance). Hence, while the existence of the analogue magnitude system and its role for quantitative cognition is well established in a large variety of species, it remains an open question whether and under which circumstances an object file system is applied in quantitative tasks, as well.

3.4. Neurobiology of quantitative cognition

3.4.1. Neural substrate of numerical cognition: IPS and PFC

A different approach to explore commonalities and differences of human and nonhuman numerical cognition is to identify and compare the neural basis of numerical competences. Neuroimaging techniques as well as electrophysiological measures have led to the conclusion that numerical processing involves similar neural substrates in humans and in nonhuman primates (see, e.g. Cantlon, 2015 for a review). More specifically, both in humans and in nonhuman primates two areas in the brain have been predominantly linked to numerical cognition: The intraparietal sulcus (IPS) within the parietal lobe, and the prefrontal cortex (PFC; see, e.g. Butterworth, 1999; Cantlon, 2015; Nieder, 2005; Nieder & Dehaene, 2009 for reviews).

For example, Nieder and Miller (2004) trained rhesus macaques to perform a visual numerosity judgment task. Electrophysiological measures showed that the proportion of numerosity-selective neurons was highest within the intraparietal sulcus (IPS). A subsequent study (Nieder, Diester, & Tudusciuc, 2006) found that, depending on whether quantities were presented simultaneously or sequentially, distinct populations of neurons along the IPS were activated. The final result of the

quantification process, however, was always coded by a third population of neurons in the IPS, suggesting that the final cardinality of a quantity is stored independently of the presentation format.

Interestingly, neurons within and around the IPS do not only represent numerosities, but also non-numerical magnitudes, such as physical size and duration (see, e.g. Cantlon et al., 2009; Kadosh, Lammertyn, & Izard, 2008), supporting the hypothesis that processing of discrete and continuous quantities share the same cognitive mechanism (see section 3.3.1.2.).

3.4.2. "Supramodal" and ratio dependent numerical tuning

Individual neurons, both in the IPS and in the PFC, seem to be "tuned" to certain numerical values: these neurons respond preferentially and at highest rates when a certain numerical value is presented. Importantly, this numerical selectivity is modality independent. For example, Nieder (2012) trained monkeys to discriminate the number of tones and light flashes while the neuronal activity in the IPS and PFC was recorded. Results revealed that populations of neurons within both regions were active during visual and auditory presentation of stimuli. Moreover, it could be shown that single neurons within the IPS and PFC are "supramodal", i.e. they respond to a certain numerical value regardless of the modality in which this numerical value is presented. However, the response of numerosity-selective neurons is imprecise: While they fire at highest rates in response to their "preferred" numerical value, they also fire in response to adjacent numerosities, and the fire rate decreases with greater numerical distance from the preferred value (the distance effect; Nieder, Freedman, & Miller, 2002; Nieder & Miller, 2003, 2004). Interestingly, neurons tuned to larger numerical values seem to respond less accurately than those tuned to smaller numbers (the magnitude effect; Nieder & Merten, 2007). Consequently, representing and processing larger quantities is noisier, and therefore less accurate than processing of smaller quantities. In combination, the distance and magnitude effect depict a neural version of Weber´s law and are therefore likely to be the cause for the ratio dependent analogue representation of number (see, e.g. Nieder, 2012). Similar findings have been reported for humans: Functional magnetic resonance imaging (fMRI), for example, enabled the detection of ratio-dependent neural responses in the IPS of human adults (Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004) and children (Cantlon, Brannon, Carter, & Pelphrey, 2006; Cantlon et al., 2009). Likewise, ratio-dependent numerical tuning was observed in pre-verbal infants using electroencephalogram recordings (EEG; Hyde & Spelke, 2011; Libertus, Pruitt, Woldorff, & Brannon, 2009). The IPS and the PFC, therefore, represent analogue numerical values both in humans and nonhuman primates, and exhibit ratio effects in their neural signatures.

Interestingly, the neuronal coding schema of proportional information (i.e. ratios of quantities) is highly reminiscent of that of absolute quantity both in humans and in nonhuman primates (see Jacob, Vallentin, & Nieder, 2012 for a review). More specifically, the same regions that process absolute quantities (the IPS and the PFC), also encode ratios of quantities (Jacob & Nieder, 2009; Vallentin & Nieder, 2008, 2010). Within these two cortical regions, single neurons were found to be tuned to certain ratios and their response rate decreased the further the distance from this preferred proportion was (Jacob & Nieder, 2009; Vallentin & Nieder, 2008, 2010). Hence, also proportional information seems to be represented in an analogue, approximate way, confirming findings of behavioral studies (Drucker et al., 2016; McCrink & Wynn, 2007).

The most obvious difference between human and nonhuman numerical cognition is that most human mathematical abilities ground in a symbolic number concept. As outlined above, comparative research has shown that nonhuman primates are, to a certain extent, able to associate Arabic numerals with the corresponding cardinal values (e.g. Boysen & Berntson, 1995; Matsuzawa, 1985). Intriguingly, neuroimaging techniques revealed that the neural substrate of symbolic representation of number is, at least partially, overlapping with that of analogue numerical processing, both in humans (see e.g. Cantlon, 2015; Dehaene et al, 2003 for reviews) and in nonhuman primates (Diester & Nieder, 2007). For example, in rhesus macaques trained to associate Arabic numerals with quantities, specific neurons in the PFC responded selectively, regardless of whether the numerical value was presented as analogue quantity or as Arabic numeral (Diester & Nieder, 2007). This suggests that human symbolic number concepts and sophisticated mathematical skills may have derived from evolutionary ancient, analogue structures (see, e.g. Cantlon, 2015 for discussion).

In sum, the neuronal basis of numerical cognition in nonhuman primates is very similar to that of humans and mainly involves two areas in the brain: the IPS and the PFC. The characteristic signatures of the analogue magnitude system are grounded in the ratio-dependent numerical selectivity of neurons in these areas. This evolutionary ancient neural substrate is most likely the foundation of the human symbolic number system.

3.5. Conclusions

Comparative research revealed fundamental commonalities in basic quantitative cognitive abilities of humans and nonhuman primates. Just like humans, great apes and monkeys (as well as many other

species) are able to represent and compare numerosities and other types of magnitudes. Moreover, they can represent proportions and perform simple arithmetic operations like summation and subtraction. To a limited extend, nonhuman primates are even able to associate Arabic numerals with the appropriate quantities. Quantity representations are subject to specific signatures following Weber´s law, which are caused by their analogue representation format. These characteristic signatures are grounded in the ratio-dependent selectivity of neurons in the IPS and PFC and can therefore also be verified on a physiological level. While the existence of an analogue magnitude system is well documented and undisputed, the role of an object file system, limited to small set sizes, is still highly debated. Similarly unclear is whether the analogue magnitude system supports all sorts of magnitude representations, or whether humans and nonhuman primates have a distinct system specifically for (discrete) numerical input. It is likely that the analogue magnitude system is the foundation of human advanced quantitative cognitive abilities, on which the later acquired symbolic number system is based. Therefore, one can conclude that "Human´s explicit knowledge of symbolic number and geometry may be our unique, abstract cognitive ability (…), but it clearly depends, in part, on a system that is widely shared by other animals" (Vallortigara, 2015, p. 59).

These findings have important implications for the possibility of finding statistical reasoning capacities in nonhuman species: A prerequisite for intuitive statistics is the ability to represent quantities, and in particular proportions of quantities. Moreover, one needs to be able to operate on these mental representations in order to form probabilistic computations. The here reviewed research showed that great apes and monkeys do possess these basic prerequisite abilities. Moreover, humans and other primates share a cognitive mechanism for these basic quantitative abilities, as well as its neuronal foundation. Together with the discovery that intuitive statistical capacities are present in pre-verbal infants, and are therefore independent of language, these findings give a crucial hint that nonhuman primates might be capable of basic intuitive statistics, as well. In the following chapter I will discuss whether nonhuman primates´ abilities (and potentially those of other animals) extend from representing absolute and relative quantities to representing and operating on probabilities.

4. Probabilistic reasoning in nonhuman animals

Do nonhuman primates´ abilities extend from representing absolute and relative quantities to representing and operating on probabilities? Are great apes, like human infants, intuitive statisticians, i.e. able to reason from population to sample and vice versa, even in contexts they have never experienced before? In the following section I will summarize findings from biological and psychological research which gave first hints that also nonhuman species might be capable of basic forms of statistical inferences. In particular, I will focus on studies investigating decision-making processes and risk-taking strategies, mainly in foraging contexts. Subsequently, I will review first direct evidence for intuitive statistical capacities in nonhuman primates and discuss their implications for the general aim of tracing the evolutionary roots of intuitive statistics.

4.1. Decision-making under uncertainty and risk evaluation

4.1.1. Ecological relevance

Similar as for humans, most problems faced by animals throughout their lifetime- foraging for food, finding mates, avoiding predators- require them to make decisions under uncertainty. This uncertainty can either occur because parts of the information are completely unavailable or ambiguous (e.g. when deciding to seek out a never visited foraging area), or because the occurrence of an event follows a probabilistic pattern (e.g. when deciding between two trees with different probabilities of yielding ripe fruit at this time of the year). Quite often these decisions have highly relevant consequences for the animal- either by directly influencing its own survival rate (e.g. in the context of foraging decisions), or by influencing its reproductive success (e.g. in the context of mate choice). Considering these direct fitness consequences, it is reasonable to assume that natural selection has favored mechanisms to cope with uncertainty. In general, there are three considerable (and not mutually exclusive) possibilities how an animal can make reasonable decisions under uncertainty: First, it can be optimally adapted to its environment and be equipped with innate decision rules which allow to use certain environmental cues in order to produce a certain behavior. For example, even plant seeds possess mechanisms to detect and combine different sensory cues, and only germinate when a photoperiod follows *after* a long period of winter coldness, because then chances are high that spring has arrived, and not just a warm weather front in autumn (Bradbeer, 1988). Second, an animal can be equipped with some form of behavioral flexibility

and an ability to learn from past experiences, and it is able to recall these learned contingencies in future situations. For example, a long-lived animal might have experienced that a certain tree is very likely to be fruiting during a particular time of the year, while the fruit of other trees tends to be ripe a little later. Hence, certain environmental cues in combination with the according physiological state (hunger) will evoke the memory of the fruiting tree and lead the animal to decide to visit this location. A third possibility is that an animal possesses a domain general inference mechanism, allowing for rational decisions in not previously encountered events by estimating probabilities.

A great deal of research suggests that most animal species possess at least option 1, often in combination with option 2, and thus have an "adaptive toolbox" of simple rules of thumb and experienced-based strategies to rapidly make decisions under uncertainty in specific environmental contexts (see, e.g. Gigerenzer, 2008; Gigerenzer, Todd, & ABC Research Group, 1999; Hammerstein & Stevens, 2012; also see Staddon, 1988 for a description of animal learning as Bayesian inference). This adaptive toolbox even allows to make "Bayesian like" decisions: Updating prior, perhaps innate "knowledge" with new experiences to optimally adjust behavior to current situations can lead to "close to optimal" decision making. Relatedly to the Bayesian approach to human statistical inferences, the "Bayesian decision theory" aims to investigate how animals integrate prior information and observations in accordance with Bayes´ theorem and whether they make decisions that are optimal given the appropriate posterior probability. To date, Bayesian computational models have been used to explain various cognitive processes of animals from simple sensorimotor reactions to more complex behaviors like foraging decisions (e.g. Biernaskie, Walker, & Gegear, 2009), mate choice (e.g. Collins, McNamara, & Ramsey, 2006) and collective behavior (e.g. shoaling; Pérez-Escudero & Polavieja, 2011). Some studies compared empirical observations of animal behavior with predictions derived from different types of decision-making models. One study, for example, investigated reproductive decision-making in the peacock wrasse (*Symphodus tinaca*; Luttbeg & Warner, 1999). During spawning, the female decides each day whether to leave her eggs with one of the abundant non-nesting males, or whether to invest time to find a rare nesting male, who will protect the fertilized eggs against predators. Hence, females have to estimate the probability of finding a nesting male each day. Luttbeg and Warner (1999) compared the behavior of the females with three types of decision models: One used a fixed search time, after which a non-nesting male was chosen, another one used fixed estimates of the probability of finding a mate that changed over the season, and a third one- the Bayesian model- assumed that females update the probability of finding a nesting male based on a prior distribution of nesting males (i.e. the probability

distribution of successfully finding a nesting male each day). The prior was initially uninformative (because of a lack of experience), but through daily updating based on relative frequencies of successes, it became more informative each day. The wrasse´s behavior was best predicted by the Bayesian model, and this model also resulted in highest fitness benefits compared to the other models. Other studies experimentally manipulated the information animals were given in order to compute prior probabilities, and then compared observed and predicted behavior. For example, Lima (1984) examined patch exploitation of downy woodpeckers (*Picoides pubescens*). The birds were presented with one of three environments. All of them contained artificial patches with 24 potential hiding locations each. Half of the patches were completely empty; the other half were provisioned with food in varying proportions of full to empty hiding locations. In one environment, all hiding locations in provisioned patches were filled with food. In a second environment, provisioned patches were only half full (i.e. 12 out of the 24 hiding locations were filled), and in the third environment they were 1/4 full. The author aimed to investigate the birds´ ability to estimate that a patch was empty. The Bayesian model predicted that in environment 1, birds should infer that all hiding locations of a patch are empty after finding a single empty sample. In environment 2, by contrast, a bird should switch to a new patch after finding three empty hiding locations. In environment 3, finally, the optimal decision maker should sample six empty hiding locations before deciding that the whole patch is empty. The birds´ behavior matched these predictions (despite considerable variance between and within individuals), suggesting that the birds learned the prior distribution of patch types and combined this prior information with current sample information in order to estimate the quality of current patches. Similar results were found for starlings (*Sturnus vulgaris*; Lima, 1985).

These and other studies (e.g testing bumblebees (Biernaskie et al., 2009) and fish (Pérez-Escudero & Polavieja, 2011)) suggest that various animal species often behave in a manner consistent with Bayesian models that assume individuals combine prior knowledge (i.e. either innate "knowledge" or previously learned knowledge of certain environments) with new sample information to update their estimates of environmental resources (see McNamara, Green, & Olsson, 2006; Valone, 2006 for reviews). This, in turn, indicates that many species are adapted to make reasonable decisions under uncertainty within a specific context. However, even though these behaviors exhibit a certain flexibility, it is plausible that animals do in fact not "compute" posterior probabilities as basis for their decisions, rather they probably use different rules of thumb/heuristics that can be flexibly adjusted depending on the current input.

What about option 3? Is it even necessary to have a domain-general statistical inference mechanism, if the "adaptive toolbox" apparently leads to good enough solutions in uncertain situations? In comparison to innate decision rules and learning mechanisms, intuitive statistical abilities bring some clear advantages: First, they are highly flexible without necessarily relying on past experiences, and hence give individuals an advantage in new situations in a changing environment requiring rational inferences. This, in turn, may open up new ecological niches, which are inaccessible for less flexible individuals. Second, as previously elucidated, intuitive statistical inference mechanisms do not only help individuals to assess event probabilities, they also constitute promising candidate mechanisms for rapidly acquiring knowledge about the world from limited sample observations. Individuals who have a genuine understanding of their physical and social environment will be able to adapt their behavior more flexibly and rationally than individuals whose knowledge is restricted to innate and directly experienced events. More generally speaking, being able to generalize from limited examples and to use the generalizations to make rational decisions in uncertain situations brings clear fitness benefits in a variety of contexts, especially for long-lived species living in changing environments.

4.1.2. Empirical evidence for sensitivity to probability

4.1.2.1. Sensitivity to probabilistic reinforcement and temporal patterns

As described above, nonhuman primates possess basic numerical abilities and can represent quantities and operate on these representations in ways comparable to humans´ non-symbolic number capacities. Quantitative processing, especially the ability to represent proportions, probably plays a crucial role in statistical inferences. But do nonhuman quantitative skills extend from representing proportions to representing probabilities?

Many behavioral studies found that various animal species exhibit a certain sensitivity to probabilistic reward patterns. For example, bees (Real, 1991), cockroaches (Longo, 1964), fish (Behrend & Bitterman, 1961), birds (Bullock & Bitterman, 1962; Graf, Bullock, & Bitterman, 1964), rats (Wodinsky, Bitterman, & Candland, 1958) and monkeys (Meyer, 1960; Wilson, Oscar Jr, & Bitterman, 1964) all seem to be able to adjust their expectations and actions in response to proportional differences in reward distribution. For example, when presented with two stimuli, one was rewarded in, e.g. 70% of trials, the other in 30% of trials, the African mouthbreeder (*Tilapia macrocephala*) learned over time to prefer the stimulus with a higher probability of being rewarded (Behrend & Bitterman, 1961). Importantly, in these

studies absolute and relative frequencies of reinforcement were usually confounded. Hence, it is not possible to state whether subjects learned to estimate the probability for each stimulus to lead to a reward, or rather just learned to pick the stimulus which was rewarded more often. Moreover, the training for these studies usually took several weeks, impeding any conclusion about intuitive abilities. Interestingly, in these probability-learning paradigms, many species (including humans under certain conditions (see Vulkan, 2002 for a review)) did not maximize their outcome by developing a clear preference for the proportion-wise more favorable option; instead, they "matched" the probabilities, i.e. when one option delivered food in 70% of trials, the other in 30% of trials, some species would choose the latter in roughly 30% of trials, instead of sticking with the former throughout the test (e.g. Behrend & Bitterman, 1961; Bullock & Bitterman, 1962; Longo, 1964). These findings may suggest that those species are unable to use probabilistic information in order to maximize their gains and instead rely on suboptimal decision rules like "win-stay, lose-shift"-strategies. However, it is likely that animals´ suboptimal behavior in these studies was due to methodological issues. For example, in the studies on fish (Behrend & Bitterman, 1961) and pigeons (Bullock & Bitterman, 1962), subjects were trained with a "guided-choice" paradigm. Thus, when picking the stimulus which was not reinforced in that particular trial, this stimulus was removed, and fish were "forced" to pick the other one. This procedure might have facilitated the usage of low-level choice heuristics. Studies using different procedures found evidence for maximizing, instead of matching strategies, for example in rats (Wodinsky et al., 1958) and monkeys (Meyer, 1960; Wilson, 1960; also see Shanks, Tunney, & McCarthy, 2002 for a study reporting how methodological changes can lead to optimal decisions in humans).

More recent studies investigated animals´ ability to choose the optimal aiming point in a probabilistic timing-task. For example, in one study (Balci, Freestone, & Gallistel, 2009) mice and humans had to anticipate at which of two locations a reward would appear. In some trials the reward appeared with a short latency at one location, in other trials it appeared after a longer latency at the other location. Subjects´ task was to integrate two factors of uncertainty: The exogenously varied probability that a given trial was either a short or a long trial, and the endogenously variable estimate of elapsed duration. More specifically, subjects began each trial waiting in front of the short-latency location. When they estimated that the short-latency time had passed without a reward being delivered, they switched to the long-latency location. Switching too soon or too late resulted in a loss of the reward. Both mice and humans accurately assessed both types of uncertainty to compute the optimal target latency for a switch. These

findings suggest that when probabilities are directly experienced, both species seem to be able to consider probabilistic information in the context of temporal decision making.

In sum, many species are able to learn to adjust to a stochastic environment and seem to be sensitive to probabilistic reinforcement and temporal patterns. However, all of the described studies required extensive training, and often confounded absolute and relative frequencies of reward distribution. Hence, from these studies alone it remains an open question whether animals other than humans can truly use probabilistic (rather than absolute frequency) information in intuitive setups to make rational decisions.

4.1.2.2. Probability-sensitive risk-taking strategies of nonhuman primates

Nonhuman primates´ intuitive sensitivity to differences in probabilities of finding food has been demonstrated in studies investigating risk-taking strategies (e.g. Haun, Nawroth, & Call, 2011; Pelé, Broihanne, Thierry, Call, & Dufour, 2014; Petrillo, Ventricelli, Ponsi, & Addessi, 2015; Proctor, Williamson, Latzman, Waal, & Brosnan, 2014). In one study (Haun et al., 2011), for example, apes (chimpanzees, bonobos, orangutans and gorillas) were allowed to choose between two options: One was a food item of medium value (a small banana piece) and depicted the safe option. The other was a high value food item (a large piece of banana), but it was hidden underneath one of several cups. When apes chose this risky option, they could point to one of the cups and received the reward only if it was the one concealing the food. The authors varied the number of cups (between one and four), resulting in different probabilities of finding the food in the risky option, as well as the size of the safe option reward. Individuals of all tested species chose the safe option more often with decreasing probability of success in the risky option. Equivalent results were found for capuchin monkeys (*Sapajus spp.*; Petrillo et al., 2015). Another study investigated nonhuman primates´ willingness to exchange a fixed medium value food item in order to gamble for a better one: Orangutans, capuchin monkeys and long-tailed macaques were presented with one medium sized piece of cookie and a tray of six cups, each containing a piece of cookie that was either larger, smaller or of the same size as the individual piece. After the subject had seen the content of the cups, it was handed the individual piece of cookie. Then it had the opportunity to give it back to the experimenter to receive the content of one randomly chosen cup. The chances of gaining or losing by exchanging the medium sized cookie piece were manipulated via different combinations of rewards in the cups (e.g. four larger pieces and two smaller pieces in one condition; one larger piece, three medium and

two smaller pieces in another). The decisions of individuals of all three species to keep their initial piece of cookie or to gamble for more were affected by the probability of winning in each trial: The higher the chances of winning a larger piece of cookie, the higher the likelihood that subjects traded their initial piece. Conclusively, when deciding between safe and risky options, nonhuman primates seem to consider the probability of success. However, other studies demonstrated that, just as for humans, also various other factors influence apes´ and monkeys´ risk taking strategies. Among these factors are social context (apes become more risk-prone after competitive events; Rosati & Hare, 2012; for similar findings in humans see, e.g. Ermer, Cosmides, & Tooby, 2008), framing (i.e. whether an option is presented as potential gain or potential loss; Krupenye, Rosati, & Hare, 2015; Lakshminarayanan, Chen, & Santos, 2011; for similar findings in humans see, e.g. Tversky & Kahneman, 1981), the frequency of losses or wins in previous trials (e.g. Pelé et al., 2014; for similar findings in humans see, e.g. Ayton & Fischer, 2004), and general species specific innate tendencies (e.g. chimpanzees seem to be more risk-prone than bonobos, potentially tracing back to differences in feeding ecology; Haun et al., 2011; Heilbronner, Rosati, Stevens, Hare, & Hauser, 2008; Rosati & Hare, 2013).

In sum, in studies investigating risk-taking strategies, subjects usually choose between one option that is stable but less preferred and one option that is variable but potentially more attractive and the main rationale is to examine which factors influence the subject´s decision and to compare these factors among humans and nonhuman species. Obviously, the objective probability of winning is one factor determining whether subjects choose the uncertain or the stable option, suggesting that nonhuman primates have some intuitive sense of probability guiding their choices. But can nonhuman species also discriminate between two uncertain options, i.e. options that both vary in their probability of yielding a reward and choose the option depicting the higher probability of leading to a preferred outcome? Does great apes´ intuitive sense for probability extend to intuitive statistical abilities, i.e. to the capacity to use sample information in order to draw conclusions about a population and vice versa?

4.2. Intuitive statistics in nonhuman primates

One recent study investigated chimpanzees´ capacity to discriminate two options that differed in their relative likelihood of finding a food reward: Hanus and Call (2014) presented chimpanzees with two trays, each holding one to six cups. The subjects witnessed that some of the cups were baited with food (up to four), then they were shuffled. Hence, subjects knew how many cups were baited on each tray, but not

which ones. Subsequently chimpanzees could pick one of the cups (from either of the two trays), and received the food reward in case it was hidden underneath that cup. From the first trial onwards, the apes´ choice depended on the ratio of food to cups on each tray and thus on the likelihood of finding food. Importantly, chimpanzees did not simply choose the tray where more food items were hidden. Instead, they took into account the ratio of food to cups on each tray. The higher the discrepancy between the two trays´ food/cup ratios, the easier it was for the subjects to choose the tray depicting the higher probability of finding hidden food. Interestingly, apes´ performance depended on the magnitude of difference between the two trays even in conditions where one of them constituted a 100% likelihood of finding food. For example, chimpanzees performed better in a condition in which a 1/2 (food to cup) ratio was contrasted with a 1/4 ratio, compared to a condition where a 2/2 ratio was contrasted with a 4/6 ratio. This suggests that the ratio between ratios, more than the magnitude of difference within the single ratios, was decisive for apes´ ability to discriminate probabilities. However, chimpanzees´ success in this study could be explained with an absolute quantity heuristic, not regarding the amount of food items available, but regarding the number of cups on each tray. In particular, the tray depicting the more favorable food/cup ratio always held the smaller number of cups. The authors added a control condition in which they excluded simple associative learning explanations (subjects did not preferentially choose the tray with fewer cups when the food was visibly removed from all cups). Nevertheless, it cannot be excluded that, as soon as there was any food to be found, chimpanzees at least partially relied on a mental shortcut such as "fewer cups= higher likelihood of finding food". Nevertheless, this study was among the first to suggest that nonhuman great apes might be able to reason about single-event probabilities and discriminate between two options depicting different probabilities of leading to a food reward.

At the starting time of the current dissertation, there was only a single study examining whether great apes are capable of intuitive statistics in ways similar to human infants. Rakoczy and colleagues (2014) modelled a paradigm after those used in infant studies (Denison & Xu, 2010a), to investigate whether apes are capable of the most basic form of intuitive statistics: the ability to reason from a population to randomly drawn samples. In a series of seven experiments, chimpanzees, gorillas, orangutans and bonobos were presented with two transparent buckets containing mixtures of preferred and non-preferred food items (banana pellets and carrot pieces) in specific ratios. Subsequently, the experimenter randomly drew one sample each in a way that the subject could not see what was drawn. Then the subject was given a choice between the two covered samples. In order to receive a preferred food item, apes thus had to discriminate between the two populations with regard to their proportions

of preferred to non-preferred food items, and form according expectations about the probability of sampling a preferred food item from each of them. Apes across conditions were able to infer which of the two populations was more likely to lead to a pellet as a sample, and this was apparent from the very first trial onwards. Moreover, they chose correctly even when absolute and relative frequencies were disentangled, i.e. when the population with the more favorable ratio of pellets to carrots contained absolutely fewer pellets than the other one. Additional control experiments ruled out alternative explanations such as "clever Hans" phenomena or usage of olfactory cues. One alternative explanation that could not be ruled out is that that apes may have applied an avoidance heuristic based on the absolute quantity of carrots. More specifically, subjects could have succeeded in all conditions by choosing the sample drawn from the population with fewer carrot pieces. Even though this alternative explanation could not be directly excluded in this study (whereas it was addressed in an infant study; see Denison & Xu, 2014), the performance patterns across experiments make it seem unlikely that apes actually relied on this mental shortcut: Subjects did not perform better in experiments in which the magnitude of difference between the absolute quantity of carrots was larger, compared to conditions in which it was smaller. Therefore, this study was the first to suggest that nonhuman great apes, just like human infants, can reason from population to sample based on proportional information. This indicates that a basic form of intuitive statistics is not uniquely human, but shared with our closest living relatives, the great apes.

4.3. Conclusions and open questions

Great apes and monkeys do not only share with humans the capacity to approximately represent absolute and relative quantities. They also possess a sensitivity towards probabilistic information, which seems to aid nonhuman primates in the context of decision-making and risk evaluation. Furthermore, first evidence demonstrated that nonhuman great apes are able to reason about never before experienced single-events and to draw intuitive inferences from population to randomly drawn samples much like human infants. These important findings are a first hint that great apes might possess intuitive statistical reasoning capacities comparable to those recently found in human infants. At the same time, these findings raise a range of fundamental questions: First, how flexible are apes´ intuitive statistics? True intuitive statistical competences require bi-directional inferences, i.e. from population to sample as much as from sample to population. Do great apes´ statistical abilities show the same flexibility as has been shown for human infants? Second, what are the cognitive structures and limits of intuitive statistics? And

are these the same both in humans and in great apes? One possibility is that intuitive statistics builds on the same, evolutionary ancient cognitive process as many other aspects of numerical cognition: the analogue magnitude system. If that was the case, characteristic signature limits should be found both in apes and in humans. Lastly, are great apes - like human infants - able to integrate intuitive statistical inferences with knowledge from other cognitive domains? Do they understand that under some circumstances drawing can be non-random? In the present dissertation I sought to illuminate these profound questions.

5. Aim of dissertation

The overall aim of this dissertation was to shed light on the evolutionary roots of intuitive statistics by exploring such capacities in humans´ closest living relatives, the great apes. Finding comparable abilities in nonhuman primates as have been found for human infants would give us important insights in both the phylogenetic origins of humans´ cognitive capacities, as well as in the cognitive architecture of great apes. Previous research demonstrated that nonhuman great apes seem to share with humans the fundamental capacity to draw statistical inferences from populations to randomly drawn samples, providing reasons to hypothesize that intuitive statistics might not be a uniquely human capacity. In the present dissertation, I sought to investigate the structure of intuitive statistical abilities in great apes more comprehensively and systematically, and, from a comparative point of view, to explore the commonalities and differences between great apes and humans. In a series of three studies I used variants of a previously established paradigm to address three distinct but related questions regarding the flexibility of great apes´ intuitive statistical abilities, their cognitive foundations and interaction with other cognitive domains.

The rationale of Study 1 was to examine whether great apes are able to reason statistically from multi-item samples to populations of food items. Research with human infants has documented flexible intuitive inferences both from population to sample (e.g. Denison & Xu, 2010a) and vice versa (Denison et al., 2013; Xu & Garcia, 2008). Although the findings of Rakoczy et al. (2014) raised the possibility that apes and human infants may operate with the same cognitive capacities for intuitive statistics, it remains an open question whether apes' intuitive statistics reveal the same kinds of flexibility and generality as those found in human infants. In particular, does apes´ ability to form expectations about samples randomly drawn from populations (inference population → sample; Rakoczy et al., 2014) extend to the ability to reason from a given sample to the corresponding population (inference sample → population)?

The main aim of Study 2 was to explore the cognitive structures and limits of intuitive statistics in apes and human adults. Previous research has demonstrated that apes, much like human infants, can draw statistical inferences from population to sample. To date it remains unclear which cognitive mechanism underlies and enables this capacity and whether it is the same in both species. One possibility is that intuitive statistics builds upon the same foundation as other quantitative abilities, the analogue magnitude system (see section 3.3.1. for further explanation). If that was the case, one would expect the characteristic performance patterns and signatures both in humans and in nonhuman great apes. Much like the discrimination of absolute set sizes varies as a function of the ratio of the set sizes, I expected that the accuracy of intuitive statistics would vary as a function of the ratio of ratios between sets. As a second

goal, I sought to replicate and validate previous findings on intuitive statistics in great apes by testing completely naïve individuals who were raised and housed in a different environment than the previously tested population. Lastly, I designed this study to comprehensively rule out alternative explanations based on absolute quantity heuristics that had not been sufficiently addressed in previous work (Rakoczy et al., 2014).

In Study 3, the guiding question was whether intuitive statistics is cognitively integrated with the processing of other types of information in great apes. Recent work in developmental psychology has demonstrated flexible integration of processing statistical information with other types of information, most notably physical and psychological information (e.g. Denison et al., 2014; Teglas et al., 2011; Xu & Denison, 2009). To investigate whether apes consider information about psychological states of experimenters (their preferences and visual access) when drawing statistical inferences, I combined the methodology used by Rakoczy et al (2014) and Xu and Denison (2009). More specifically, I varied whether the sampling experimenters had preferences to draw certain types of food or were neutral and whether they had visual access to the population or sampled blindly. If chimpanzees, like human infants, were able to consider the experimenters´ mental states in the context of statistical reasoning, we would expect them to adjust their choices depending on whether drawing is random (and therefore the outcome is determined by the proportional composition of the population), or non-random (and therefore the outcome is determined by the experimenters´ intentions).

6. Summary of empirical findings

In the following section I will summarize the main findings of the three studies I conducted in the course of this dissertation: (1) Eckert, Rakoczy and Call (2017) – Are great apes able to reason from multi-item samples to populations of food items?; (2) Eckert, Call, Hermes, Herrmann and Rakoczy (2018) – Intuitive statistical inferences in chimpanzees and humans follow Weber's Law; (3) Eckert, Rakoczy, Call, Herrmann and Hanus (2018) – Chimpanzees consider humans´ psychological states when drawing statistical inferences. For further details regarding subjects, design, procedure, analysis, and results, please refer to the original manuscripts (Appendices A, B, and C).

6.1. Eckert, Rakoczy, and Call (2017)

Previous work has shown that great apes are able to draw statistical inferences from population to sample (Rakoczy et al., 2014). The aim of this study was to investigate whether apes possess similar capacities in the other direction: From sample to population. To address this question, I tested 26 zoo-housed great apes (ten chimpanzees, six bonobos, six orangutans, and four gorillas) in a task that combined the methodology of Rakoczy et al. (2014) and Xu & Garcia (2008). More specifically, I confronted apes with two covered containers holding populations of food items that differed in their proportion of preferred to neutral food. In two experiments, the apes witnessed the experimenter drawing a multi-item sample from each population. The distribution within the samples reflected the distribution of the respective populations. Based on the observation of these representative samples, the apes were given the choice between the two covered populations. Hence, to receive the more favorable population, they were required to use the proportional information provided by the samples. In Experiment 1 apes were confronted with two populations (24:6 vs. 6:24) and witnessed the samples 4:1 versus 1:4 being drawn from these. Two control conditions were designed to rule out alternative explanations: In order to test whether the information provided by the samples alone was sufficient for the apes to infer about the distribution within the populations, apes did not see the available populations beforehand in Control 1. In Control 2, the samples were not re-inserted into the populations after the drawing process. This manipulation tested whether apes' success in the previous conditions might have reflected a tendency to choose the population where the more favorable sample was inserted, without necessarily having to reason about the drawing process. In Experiment 2, apes were tested in two further test conditions in which absolute and relative frequencies of preferred food items within samples were disentangled: In the

2:1 versus 4:8 Test, the absolute number of preferred food items was lower in the sample drawn from the more favorable population and therefore misleading. In the 4:1 versus 4:8 Test condition, the absolute number of preferred food items was the same in both samples and therefore inconclusive. Hence, to receive the more favorable population in Experiment 2, apes had to take into account proportions, rather than absolute numbers of preferred food items.

Results showed that apes seemed to make inferences from samples to populations and thus chose the population from which the more favorable (4:1) sample was drawn in Experiment 1, irrespective of whether they knew the composition of the available populations beforehand or not (Control 1) and whether samples were replaced after drawing or not (Control 2). The results of Control 2 are especially revealing, as they rule out the possibility of a simple heuristic: "choose the container where the more attractive sample was inserted". Instead, apes apparently considered the drawing process and reasoned about the population as a whole from the first trial onwards. This implies that apes possess similar kinds of capacities as human infants (Denison et al. 2013; Xu & Garcia 2008).

However, based on Experiment 1 alone it was impossible to rule out that apes used alternative strategies based on the absolute number of preferred food items. The aim of Experiment 2, therefore, was to investigate whether great apes can successfully reason from samples to populations when prevented from relying on absolute quantities. Apes performed at chance level both when the sample drawn from the more favorable population contained fewer preferred food items than the other one, and when both samples contained the same number of preferred food items. Conclusively, apes did not rely on inferences from samples to populations in this experiment. There are at least two interpretations for these findings. One interpretation is that apes' failure in Experiment 2 reflects true limitations of their cognitive competences and therefore suggests that nonhuman primates' statistical abilities could be unidirectional. This would question whether apes have a true understanding of drawing processes and the relation between populations and samples. A different and perhaps more plausible interpretation for the negative findings of Experiment 2 is that they may merely reflect performance limitations imposed by the task's accessory cognitive demands. Such demands could be, e.g. the memory component and problems with inhibitory control (i.e. problems to inhibit choosing the population where absolutely more preferred food items were drawn from). An additional crucial factor that potentially masked apes´ competences is that the magnitude of difference between samples in the critical "disentangling" conditions was relatively small, perhaps even beyond the signature limits of apes´ capacities.

Thus, while the present study demonstrated that apes, like human infants (Denison et al., 2013; Xu & Garcia, 2008), could reason from sample to population when absolute and relative frequencies were confounded, it remains unclear whether they did so based on proportional information or rather by means of general-purpose heuristics.

6.2. Eckert, Call, Hermes, Herrmann, and Rakoczy (2018)

Nonhuman great apes share with humans the fundamental ability to draw statistical inferences from population to sample (Rakoczy et al., 2014) and (to a certain extent) vice versa (Eckert et al., 2017). Yet, it still remains an open question what the cognitive foundations of intuitive statistics are and whether they are the same in humans and their closest living relatives. We[5] hypothesized that intuitive statistics builds on the same foundation as other quantitative abilities: the analogue magnitude system. As described above (see section 3.3.1.), the analogue magnitude system is an evolutionary ancient cognitive mechanism which enables estimating and comparing the numerosity of arbitrary large quantities in a fast but only approximate way, with signatures following Weber´s Law (e.g. Nieder & Dehaene, 2009). Developmental and comparative research have demonstrated that tracking and comparing ratios - a prerequisite capacity for statistical reasoning - is subject to the same signatures as tracking absolute frequencies (Drucker et al., 2016; McCrink & Wynn, 2007; Vallentin & Nieder, 2008), strengthening the hypothesis that intuitive statistics may be based on the analogue magnitude system, as well. However, there is also evidence suggesting that absolute set sizes, rather than ratios, influence human infants´ performance in intuitive statistical reasoning tasks (Teglas et al., 2015), indicating that the analogue magnitude system may not be the primary cognitive mechanism enabling intuitive statistics.

The aim of the present study was to investigate the cognitive foundation of intuitive statistics in chimpanzees and humans by testing its signatures. If intuitive statistics was based on the same evolutionary ancient analogue magnitude system, characteristic performance patterns and signatures-following Weber´s law- would be expected in both species. I tested 24 sanctuary-living chimpanzees in a previously established paradigm (Rakoczy et al., 2014) which required them to reason from populations of food items with different ratios of preferred and non-preferred items to randomly drawn samples. In a series of eight test conditions, the ratio between the two ratios to be discriminated (ROR) was systematically varied ranging from 1 (same proportions in both populations) to 16 (high magnitude of

---

[5] Whenever I use the terms "we" or "us" in this or the following sections, I refer to the co-authors of the respective publication.

difference between populations). Control conditions ruled out the usage of simple choice heuristics. One hundred and forty-four human adults were tested in a computerized version of the same task.

The main result was that both chimpanzee and human performance varied as a function of the log(ROR) and thus followed Weber´s law. This suggests that intuitive statistical reasoning relies on the same evolutionary ancient cognitive mechanism that is used for comparing absolute quantities, namely the analogue magnitude system. Intriguingly, and although methods for both species were somewhat different (e.g. live demonstration for chimpanzees versus computer setup for humans) chimpanzees and humans displayed the same approximate limit: While our model predicted above chance level performance for ratios that differed by a factor of 4, it predicted failure for those that differed by a factor ≤2, thereby further strengthening the view that both species may operate with the same cognitive mechanism for intuitive statistics. In addition, our results replicated previous findings on statistical reasoning in great apes with a new population of chimpanzees with a different housing and rearing background, suggesting that intuitive statistics is in fact a common capacity in chimpanzees. Lastly, this was the first study on intuitive statistics in great apes controlling for absolute number heuristics both regarding preferred and non-preferred items as well as total amount, providing further evidence for true intuitive statistical reasoning in chimpanzees.

6.3. Eckert, Rakoczy, Call, Herrmann, and Hanus (2018)

One important characteristic of human statistical reasoning is its functional integration with knowledge from other cognitive domains from early infancy onwards (Denison et al., 2010b; 2014; Teglas et al. 2007, 2011; Xu & Denison, 2009; see section 2.4.3.). For example, infants understand that a physical constraint can turn a sampling process into a non-random event (Denison et al., 2014; Teglas et al., 2007, 2011). Similarly, infants consider knowledge about the mental state of an experimenter (her preferences and visual access) when drawing statistical inferences and appreciate that an intentionally drawn sample does not necessarily reflect the population´s proportions (Xu & Denison, 2009). To date, nothing is known about such cross-domain integration in any nonhuman animal, leaving some uncertainty about the origins of human statistical abilities.

The aim of this study was to investigate whether chimpanzees consider knowledge about the experimenters´ mental states (their preferences and visual access) when drawing statistical inferences. I tested 21 sanctuary-living chimpanzees in a previously established paradigm (Eckert, Call, et al., 2018; Rakoczy et al., 2014) that required subjects to infer which of two mixed populations of preferred and non-

preferred food items was more likely to lead to a desired outcome for the subject. In a series of three experiments I manipulated whether experimenters had preferences to draw certain objects or acted neutrally and whether they had visual access to the population while sampling or drew blindly.

Results suggested that chimpanzees, without any prior information, assumed random sampling and expected the sample to reflect the population´s distribution. Accordingly, in this experiment they chose the sample drawn from the proportion wise more favorable population. If the apes, however, had reason to assume that the experimenters were biased, subjects´ choice reflected these biases; the severity of this influence was dependent on whether the experimenters had visual access to the population or not. More specifically, when biased experimenters had visual access to the populations while sampling, chimpanzees chose the sample drawn by the experimenter who had the same preference as they themselves, regardless of the proportional composition of the populations. When the same biased experimenters drew blindly, by contrast, subjects disregarded the experimenters´ preferences and chose based on the populations´ proportions. Importantly, chimpanzees drew inferences about the experimenter and the sampling process from the given statistical information even without being differentially rewarded: when samples were unambiguously non-representative of a populations´ distribution and the experimenter looked into the population while sampling, subjects seemed to infer that the sampling person must have a preference for one of the food types and act accordingly in the test condition.

While previous studies have shown that chimpanzees can reason probabilistically from population to sample (Eckert, Call et al., 2018; Rakoczy et al., 2014) and are sensitive to what others (both conspecifics (e.g. Hare, Call, Agnetta, & Tomasello, 2000) and human experimenters (e.g. Melis, Call, & Tomasello, 2006)) can and cannot see, this study is the first to suggest that chimpanzees are able to flexibly combine these two sources of information to make rational decisions under uncertainty. Our results resemble findings on human infants: Just as the chimpanzees in our study, 11-month-old infants were shown to be sensitive to whether a sample was drawn randomly from a population or not on the basis of information about the drawing agent's psychological states (her preference and visual access; Xu & Denison, 2009). Similar to our apes, infants were also able to use statistical information (in particular a violation of likelihoods), to draw conclusions about the sampling agent and the sampling process (Gweon et al., 2012; Wellman et al., 2016). Our study therefore gives further reason to assume that human statistical reasoning might be based on a cognitive mechanism that is utilized from early infancy onwards and shared with our closest living relatives.

7. **General discussion**

7.1. Main findings in a comparative framework

7.1.1. Generality and flexibility of intuitive statistics in great apes and human infants

The findings of the present dissertation (Eckert, Call, et al., 2018; Eckert, Rakoczy, et al., 2018) comprehensively replicate and confirm previous research (Rakoczy et al., 2014) in suggesting that great apes can draw inferences from populations to randomly drawn samples in similar ways as human infants (e.g. Denison & Xu, 2010a, 2014; Teglas et al., 2007, 2015; Xu & Garcia, 2008). Both infants and apes succeeded in active-choice tasks (apes: Eckert, Call, et al., 2018; Eckert, Rakoczy, et al., 2018; Rakoczy et al., 2014; infants: Denison & Xu, 2010a, 2014), whereas infants have additionally been tested in VOE looking time paradigms (Teglas et al., 2007, 2015; Xu & Garcia, 2008). Importantly, control conditions ruled out that subjects were relying on simple heuristics based on the absolute quantity of preferred items (apes: Eckert, Call, et al., 2018; Rakoczy et al., 2014; infants: Denison & Xu, 2014), the absolute quantity of non-preferred items (i.e. an avoidance heuristic; apes: Eckert, Call, et al., 2018; infants: Denison & Xu, 2014), or a representativeness heuristic (apes: Eckert, Rakoczy, et al., 2018; infants: Teglas et al., 2007; Xu & Denison, 2009). Instead, both human and nonhuman subjects reasoned about the proportional composition of the populations in order to assess which of them was more likely to lead to a preferred outcome. Hence, great apes can intuitively draw statistical inferences from population to sample in situations they have never experienced before. Importantly, these findings seem to reflect a general capacity in great apes: While a first study (Rakoczy et al., 2014) tested great apes who were born in captivity, mainly mother-reared, and zoo-housed, two of the present studies (Eckert, Call, et al., 2018; Eckert, Rakoczy, et al., 2018) tested chimpanzees who were born in the wild, mainly human-reared, and sanctuary-housed. Despite these different housing and rearing backgrounds, apes exhibited very similar performance patterns, indicating that the capacity to draw statistical inferences from population to sample is in fact a common capacity in chimpanzees (and other great ape species), and not dependent on certain prior experiences. Therefore, the present dissertation confirms the finding that great apes possess the capacity to draw intuitive inferences from population to sample on similar levels as human infants by the age of ten months[6].

---

[6] Infants as young as eight months succeeded in looking-time paradigms requiring inferences from population to sample (Xu & Garcia, 2008). The youngest infants successfully tested in an active-choice paradigm- comparable to that used for great apes- were ten months old (Denison & Xu, 2014).

Developmental research found that infants also possess the ability to draw inferences in the other direction: from sample to population (Denison et al., 2012; Xu & Garcia, 2008). The main aim of the first study of this dissertation (Eckert et al., 2017) was to investigate whether great apes´ statistical abilities exhibit a similar flexibility and generality. Our findings showed that, just like infants, individuals of all tested great ape species were able to use the information provided by the samples to identify the favorable population. This worked even when they did not see the populations beforehand, and, importantly, also when samples were discarded after drawing, i.e. when apes could not just track that population in which the more favorable sample was inserted. However, apes failed to choose the favorable population when absolute and relative frequencies were disentangled, i.e. when the sample drawn from the favorable population contained absolutely fewer or the same number of preferred food items that the other sample. At first glance, this may lead to the conclusion that great apes´ intuitive statistical abilities are more restricted than those found in human infants. However, a closer consideration of the testing paradigms and procedures reveals some important differences between studies on infants and great apes.

First, infants were never tested in conditions in which absolute and relative frequencies were disentangled. In both developmental studies (Denison et al., 2012; Xu & Garcia, 2008), infants could have either reacted based on a proportional rule (e.g. "it is surprising to see a sample with more red than yellow balls being drawn from a population containing more yellow than red balls") or on an absolute quantity rule (e.g. "it is surprising to see a sample containing only a few red balls being drawn from a population containing so many red balls"). The authors ruled out simple perceptual explanations to establish that infants were truly reacting towards the sampling relation between population and sample and not on the basis of a perceptual mismatch between the two sets. However, it is unclear whether infants would have succeeded in an appropriate control condition for absolute quantity heuristics. A second crucial difference between the present study and developmental studies was that, so far, infants have only been tested in looking time paradigms, while apes were tested in an active-choice paradigm. There is some evidence that findings of studies using the VOE looking time paradigm dissociate from findings of studies using active-choice measures (e.g. Ahmed & Ruffman, 1998; Balci et al., 2009; Shinskey & Munakata, 2005). This is probably due to the fact that an individual that is able to perceive something is not necessarily able to act accordingly. Therefore, it remains an open question whether infants would be able to reason from sample to population in an active-choice paradigm. The different paradigms led to further, more far reaching discrepancies: While infants looked at neutral scenes of distinctly colored Ping-Pong balls, apes reasoned

about food items. It is known from previous work (Boysen & Berntson, 1995; Ebel & Call, 2018; Schmitt & Fischer, 2011) that using food items as stimuli can add an additional level of difficulty to a task for nonhuman primates. In the "disentangled" conditions of the present study, this might have been especially challenging for the apes, since they had to inhibit the intuitive response to reach for the side where more preferred food items were drawn from. Since our aim in the present study was to investigate an intuitive capacity which is present without any prior training, we opted for using an active-choice paradigm involving food items as stimuli over arbitrary stimuli. Future studies, however, may try to use more implicit measures, such as eye-tracking, in order to achieve greater comparability between human and nonhuman studies. The application of such implicit measures would also eliminate another extraneous task demand faced by the apes in Eckert et al.´s (2017) study - the memory component: While infants could directly compare the samples and populations (Denison et al., 2012; Xu & Garcia, 2008), apes had to memorize which sample came from which population when making their decision, thereby adding yet another level of complexity to the task. A recent study confirmed that indeed both inhibitory control problems as well as memory demands severely limit apes´ performance in tasks involving inferences from samples to populations (Eckert et al., unpublished), giving further reason to assume that apes´ competences may have been masked in Eckert et al.´s (2017) study. Lastly, an important and perhaps determining factor limiting the apes´ performance in the crucial disentangling conditions was the small magnitude of difference between the samples´ ratios, which may have been well below the apes´ signature limits for intuitive statistical reasoning (see 7.1.2. for further discussion of signatures of intuitive statistics). For infants, looking times were compared between a 4:1 and a 1:4 sample (ratio of ratios, ROR=16), whereas apes had to compare a 2:1 and a 4:8 sample (ROR=4), depicting a four times smaller difference.

In sum, the work of the present dissertation confirmed and added on previous findings demonstrating that great apes share with human infants the intuitive capacity to draw inferences from population to sample. This capacity does not seem to be biased by heuristics and is independent of rearing and housing background. The evidence for commonalities in the reverse ability (drawing inferences from sample to population) still remains ambiguous. In order to draw meaningful conclusions, future studies will have to explore great apes´ abilities in tasks eliminating extraneous factors such as inhibitory control and working memory demands as well as insufficient magnitudes of differences between the samples´ proportions. Moreover, prospective research should test infants´ abilities in active-choice paradigms in which absolute and relative frequencies are disentangled.

7.1.2. Cognitive foundation of intuitive statistics in great apes and humans

The present dissertation revealed some fundamental commonalities in intuitive statistical capacities of nonhuman great apes and humans. Eckert, Call, et al. (2018) additionally demonstrated that both chimpanzees and humans may operate with the same cognitive mechanism when drawing statistical inferences: the analogue magnitude system. When presented with a task requiring inferences from population to sample, the performance of both chimpanzees and human adults was, in accordance with Weber´s Law, strongly dependent on the ROR, the relative difference between the two proportions to be discriminated. Equivalent results have been obtained in the context of absolute quantity discrimination, where (human and nonhuman) performance was a function of the ratio between the quantities to be discriminated (e.g. Beran, 2001; Beran, 2012; Cantlon & Brannon, 2007b; Hanus & Call, 2007; Izard et al., 2009), and in the context of ratio discrimination where performance varied as a function of the ratio between ratios (Drucker et al., 2015; McCrink & Wynn, 2007; Vallentin and Nieder, 2008). Also, recent work found similar ratio dependence in chimpanzees´ performance in the context of probabilistic reasoning: Hanus and Call (2014) showed that chimpanzees´ capacity to discriminate two options which differed in their relative likelihood of finding a food reward was dependent on the relative difference between the two probabilities. Hence, together with previous studies, the findings of Eckert, Call, et al. (2018) suggests that all these abilities - discriminating absolute quantities, discriminating ratios and reasoning about ratios in order to assess probabilities - may share a cognitive mechanism in humans and in nonhuman great apes, namely the analogue magnitude system.

Interestingly, also the approximate limit of this capacity was similar for both chimpanzees and humans: While our statistical model predicted above chance level performance for ratios that differed by a factor of 4, it predicted failure for those that differed by a factor ≤2. Given that this is the first study addressing this topic and considering the small sample size, these results should, however, be treated with caution. Nonetheless, the rather high threshold may suggest that, in comparison to simple quantity discrimination tasks, the statistical operation adds some error to the representation. Human adults, for instance, are able to discriminate absolute set sizes that differ by a factor of 1.15 (e.g. Barth, Kanwisher, & Spelke, 2003; Pica, Lemer, Izard, & Dehaene, 2004) compared to 6-month-old infants who can discriminate ratios >1.5 (Lipton & Spelke, 2003; Xu & Spelke, 2000). Not many studies documented the threshold for absolute quantity discrimination in nonhuman primates. Reported limits range from values as low as 0.9 for great apes (Hanus and Call, 2007) to 1.25 for rhesus macaques (Brannon & Terrace, 2000). The finding of a higher threshold for ratio discrimination within the realms of statistical inferences relative

to basic quantity discrimination is consistent with the idea that additive error is to be expected when an organism represents and operates over multiple amounts (see, e.g. Barth et al., 2006 and McCrink & Wynn, 2007 for discussion of this hypothesis).

One interesting question arising from the data of Eckert, Call, et al. (2018) concerns the more specific nature of the representation in both species. For humans, the relation between performance and ROR was clearly logarithmic. For chimpanzees, by contrast, it could be equally well described as logarithmic or linear relationship. This is related to the debate on whether the scaling that occurs in quantitative representation through the analogue magnitude system is generally linear, with increasing variability for larger magnitudes (e.g. Brannon et al., 2001; Gibbon & Church, 1981; Whalen et al., 1999), or whether it is logarithmic, with a logarithmic compression of the mental number line which is used for numerical representations (e.g. Dehaene, 2003; Roberts, 2006, also see Cantlon et al., 2009; Dehaene et al., 2008 for discussion). In the case of the present study, the finding that chimpanzees´, but not humans´, performance could have been described as a linear function might be explained by differences in accuracy: While human performance rapidly increased with higher RORs and reached a plateau at a ROR of 8, ape performance was still far from ceiling even at the highest tested ROR (16). Hence, it is possible that one would only find a clearly logarithmic curve for chimpanzees when including even higher RORs which are even easier to be discriminated. The supposition that great ape intuitive statistics might, despite a similar threshold, be less accurate than human intuitive statistics resembles findings on absolute quantity discrimination, where humans showed higher accuracy compared to other primates (e.g. Cantlon & Brannon, 2007b). Future studies should test chimpanzees with a wider range of RORs to investigate whether (and at what point) they, like humans, also reach a maximum performance plateau. Additionally, it will be interesting to test human children and infants in order to explore the developmental trajectory of the signatures of intuitive statistical abilities.

7.1.3. Integration of intuitive statistics with knowledge from other cognitive domains in great apes and human infants

Eckert, Rakoczy, et al. (2018) showed that chimpanzees do have a random sampling assumption and recognize when this assumption is violated by a biased experimenter who has visual access to the population while sampling (but not when she is drawing blindly). Very similar findings were obtained in a developmental study using the VOE looking time paradigm with 11-month-old infants (Denison & Xu,

2009). More specifically, infants looked longer at a sampling scene when the sample did not match the explicitly expressed preferences of an experimenter, but they only did so when she was looking into the bucket while sampling, and not when she was blindfolded. Chimpanzees in Eckert, Rakoczy, et al. (2018), furthermore, were able to use a violation of statistical likelihoods as indication for biased sampling: When an experimenter drew only peanuts out of a population containing mainly carrots, chimpanzees inferred that the experimenter must have drawn intentionally. They did not draw similar conclusions when she drew the same samples out of a population containing mainly peanuts. Again, very similar results were found in a developmental study applying the VOE looking time paradigm with 10-month-old infants (Wellman et al., 2016). Hence, both chimpanzees and human infants possess the ability to use information from the social domain to judge whether a drawing process is random or non-random and intuitively draw appropriate rational inferences. The fact that chimpanzees and human infants were able to infer an experimenter´s intention based on samples being non-representative of a population´s distribution also sheds more light on apes´ and infants´ genuine understanding of the relationship between populations, sampling processes and samples. More specifically, if subjects were not able to reason from population to sample *and* vice versa (as results obtained by Eckert et al., 2017 may suggest; see section 7.1.1.), they shouldn´t have been able to draw conclusions that were based on both the identity of a population *and* the sample. Finding that chimpanzees, as well as infants, could assess the consistency between sample and population gives further reason to assume that they are in fact capable to draw inferences from sample to population.

Developmental research is already one step ahead and found equivalent abilities for the physical domain: Infants seem to understand that, for instance, a cohesion-constraint or a physical barrier can prevent some objects from being drawn, resulting in a non-random event (Denison & Xu, 2010b; Denison et al., 2014; Lawson & Rakison, 2013; Teglas & Bonatti, 2016; Teglas et al., 2007, 2011). Moreover, infants are truly able to *integrate* two sources of information to compute event probabilities, instead of just flexibly deciding which of them is relevant in a particular case. For example, infants showed graded looking times, depending on the interaction of a temporal component, a spatial component and a proportional component (Teglas et al., 2011). To date there are no published studies testing for such abilities in great apes. However, preliminary work (Eckert et al., unpublished) cautiously suggests that chimpanzees might not be able to integrate information about spatial composition into their statistical inferences.

In conclusion, the present dissertation points out important commonalities between chimpanzees and human infants in their capacity to consider psychological information when drawing statistical inferences. Similar capacities in the physical domain, however, are yet to be shown in great apes.

7.1.4. Summary of comparison of great apes´ and infants´ statistical abilities

Great apes and human infants share the capacity to draw flexible inferences from population to sample. For great apes, however, inferences in the other direction might be limited: While they succeeded in conditions in which absolute and relative frequencies were confounded, they failed in conditions in which they were disentangled. Human infants, by contrast, were shown to be able to discriminate between likely and unlikely events requiring inferences from sample to population in VOE looking time paradigms. However, it remains unclear whether infants would be similarly competent in an active-choice task including crucial controls for absolute versus relative frequencies. Hence, while the capacity to draw intuitive inferences from population to sample is well established, the reverse ability still needs more exploration both in great apes and in infants. Much like human infants, great apes also take the sampling process into account when drawing inferences from population to sample and consider information of the social domain in order to judge whether a sample will reflect the population´s distribution. Relatedly, when confronted with samples which are non-representative of the population, chimpanzees (as well as infants) are able to draw conclusions about a biased sampling process. The cognitive foundation of these abilities is most likely an evolutionary ancient cognitive mechanism, the analogue magnitude system. All in all, the three studies together with previous data (Rakoczy et al., 2014) show that great apes possess similar intuitive statistical abilities as human infants (see Figure 2 for a schematic overview depicting great apes´ and infants´ intuitive statistical abilities).
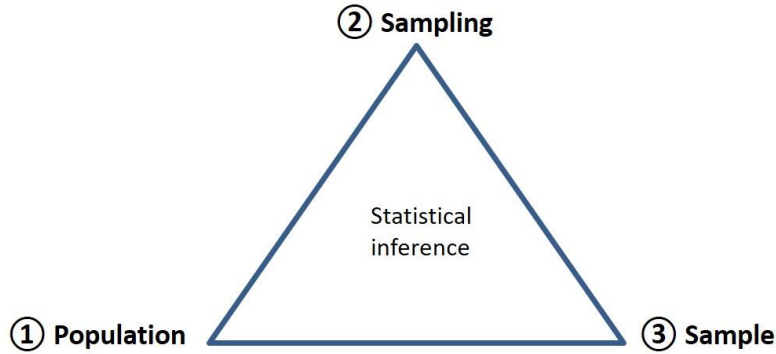
Human infants:

① + ② → ③ √ _{Denison & Xu, 2010a, 2014; Lawson & Rakison, 2013; Teglas & Bonatti, 2016; Teglas et al., 2007, 2011, 2015; Xu & Denison, 2009; Xu & Garcia, 2008}

② + ③ → ① √[?] Denison et al., 2012; Xu & Garcia, 2008

① + ③ → ② √ Gweon et al., 2010; Xu & Denison, 2009; Wellman et al., 2016

Great apes:

① + ② → ③ √ Eckert, Call, et al., 2018, Eckert, Rakoczy, et al., 2018; Rakoczy et al., 2014

② + ③ → ① ? Eckert et al., 2017

① + ③ → ② √ Eckert, Rakoczy, et al., 2018

Figure 2 Schematic overview depicting the three possible directions of statistical inferences and evidence for their occurrence in human infants and nonhuman great apes.

The introduction of this dissertation was opened with a quote of Piaget and Inhelder, highlighting the long history of research on humans´ capacity to reason about probabilities: "Could there be in a normal man an intuition of probability just as fundamental and just as frequently used as, say, the intuition of whole numbers?" (Piaget & Inhelder, 1975, xiii). Piaget and Inhelder partly answered this question with the conclusion that "there are still two perfectly normal psychological states in which the understanding of chance and probability seem more or less absent: the primitive mind and the mind of a small child" (Piaget and Inhelder, 1975, xiv). The work of the present dissertation in combination with recent developmental research as well as work on pre-literate and pre-numerate indigenous Mayan groups (Fontanari, Gonzalez, Vallortigara, & Girotto, 2014) provide evidence that this conclusion was premature. In the contrary, both great apes and human infants (as well as pre-literate indigenous tribes) possess an intuitive statistical inference device, which does not rely on past experiences and does not seem to be

biased by heuristics. Hence, the capacity of intuitive statistics clearly antedates mathematical education and even language both phylogenetically and ontogenetically.

7.2. Implications and conclusions

7.2.1. Implications for the evolutionary origins of intuitive statistics

The finding that great apes and human infants share fundamental commonalities in their intuitive statistical capacities sheds some light on the evolutionary origins of this type of reasoning. According to genetic analysis, the family of the great apes (Hominidae) emerged around 14 million years ago (Hara, Imanishi, & Satta, 2012) and comprises humans, chimpanzees and bonobos, two species of gorillas and three species of orangutans (Groves, 2005; Nater et al., 2017). Chimpanzees and bonobos are humans´ closest living relatives and most likely split from the shared lineage with humans only around 5 to 7 million years ago (Adachi & Hasegawa, 1995; Glazko & Nei, 2003; Hara et al., 2012)[7]. The work of the present dissertation suggests that chimpanzees (and most likely other great ape species: Rakoczy et al., 2014) share with humans a cognitive mechanism for basic intuitive statistical inferences. Thus, this mechanism most likely emerged before the divergence of humans and chimpanzees, i.e. at least 5 million years ago, or perhaps even earlier in evolutionary history.

Interestingly, recent studies tested both New world and Old world monkey species in tasks requiring inferences from populations of food items to randomly drawn samples, comparable to those conducted with great apes (Eckert, Call, et al., 2018; Rakoczy et al., 2014). Placi et al. (submitted) found that long-tailed macaques, at the group level, succeeded only in conditions in which they could rely on a quantity heuristic dealing with the absolute number of preferred food items (e.g. 64:16 vs. 16:64 preferred to non-preferred food items). When absolute and relative frequencies were disentangled (48:12 vs. 12:192), the monkeys failed the task, suggesting that they were not able to draw statistical inferences based on proportional information. Capuchin monkeys, by contrast, succeeded even in a crucial condition in which absolute and relative frequencies were disentangled (Tecwyn, Denison, Messer, & Buchsbaum, 2016). However, their performance in a non-preferred food avoidance control was not significantly

---

[7] Please note that these dates are still debated. Most likely, speciation was a lengthy process including periods of hybridization (e.g. Patterson, Richter, Gnerre, Lander, & Reich, 2006), making it difficult to determine exact dates.

different from chance level[8] (whereas chimpanzees were successful in a comparable condition in Eckert, Call, et al., 2018), leaving some uncertainty as to whether their success in other conditions may trace back to the usage of an avoidance heuristic. Therefore, to date there is no unambiguous evidence for the presence of intuitive statistical abilities in monkeys equivalent to those of great apes and human infants. Although much more research is needed, this may suggest that statistical reasoning emerged rather late in primate evolution, perhaps only in the ape lineage.

However, very recent research on two bird species points out that analogue abilities may have evolved in more distantly related animal classes. One study (Clements, Gray, Gross, & Pepperberg, 2018) presented language-trained grey parrot Griffin with a set of four objects, consisting of a 3:1 ratio of two item types (e.g. corks and keys). The experimenter placed the objects in an opaque container. Subsequently, she randomly drew one item in a way that the subject could not see what was drawn, and then asked the subject about the hidden object´s identity. The parrot vocally responded in roughly 75% of trials that the object was of the majority type, thereby matching the proportional distribution of items. While this study demonstrated a certain sensitivity to probabilistic patterns, it is important to note that the procedure differed from those applied in primate studies in several ways: First, training for this study (not including the language training), lasted several months. The grey parrot was trained in several steps until he reached criterion as indication of his understanding of the drawing-procedure. Relatedly, the testing phase (for a single ratio) lasted 96 trials; it is, therefore, controversial whether the parrot exhibited an intuitive capacity or rather a stepwise trained response. Furthermore, during these 96 test trials the parrot was rewarded when he guessed the drawn object correctly, resulting in a probabilistic reward pattern reflecting the population´s distribution. This resembles early studies where sensitivity to probabilistic reward patterns was found in various species ranging from bees (Real, 1991), cockroaches (Longo, 1964), and fish (Behrend & Bitterman, 1961), to rats (Wodinsky et al., 1958) and monkeys (Meyer, 1960; Wilson, 1960). Similar as the grey parrot in Clements et al. (2018), many other species exhibited "probability matching" instead of maximizing their outcome (e.g. picking a 70% rewarded stimulus in roughly 70% of trials, instead of 100%; e.g. Behrend & Bitterman, 1961; Bullock & Bitterman, 1962; Longo, 1964). Hence, the study design of Clements et al. (2018) leaves some uncertainty as to whether grey

---

[8] The authors reported that the monkeys´ performance in the avoidance control was marginally significant (P=0.063). However, results only reached marginal significance after the data of all those subjects, who showed a side bias in this condition, were removed. Since exhibiting a side bias can be a valid strategy for subjects who do not understand the task, removing the data of those individuals leads to an exclusion of unsuccessful subjects, and therefore to a systematic bias towards (non-representative) positive results.

parrots can truly use proportional information within a population to draw *intuitive* inferences about randomly drawn samples, or whether they simply possess a certain sensitivity towards probabilistic reward patterns. Another recent study showed that pigeons are sensitive to probabilistic reward patterns, independent of the absolute quantity of rewards (Roberts, MacDonald, & Lo, 2018). More specifically, pigeons learned that each of two individually presented keys would lead to the delivery of a food reward in a certain proportion of trials when pecked on. When subjects had the choice between the two keys, they strongly preferred a 75% rewarded key over a 25% rewarded key and thus showed clear signs of maximizing as opposed to suboptimal probability matching. Importantly, they even preferred the proportionally favorable key when the total number of rewards obtained on both keys was equated. Hence, pigeons are able to extract proportional information from past experiences in order to adapt their behavior in the most beneficial way.

Although the studies conducted with birds so far are not directly comparable with those conducted with human infants and nonhuman primates, they do suggest that birds, too, have a sense of probability. To date, it remains an open question whether this sense of probability is a product of an intuitive statistical inference mechanism, and whether birds, like primates, are able to intuitively grasp the relation between populations and samples without relying on past experiences. However, considering that the analogue magnitude system, which might be the foundation of intuitive statistics, is present in a wide range of species and that the ability to draw intuitive statistical inferences most likely brings significant fitness benefits, the idea that intuitive statistics may have emerged several times during evolution in a convergent manner does not seem too far-fetched.

In sum, even though much more research is needed to draw valid conclusions, the evidence to date suggests that intuitive statistics might have evolved in the last common ancestor of humans and nonhuman great apes. An analogue capacity may have evolved independently in other species, as suggested by recent studies on two bird species (Clements et al., 2018; Roberts et al., 2018).

7.2.2. Implications for the cognitive architecture of great apes

The present dissertation´s findings give important insights into the workings and architecture of the mind of great apes. Chimpanzees and other species of great apes are evidently able to intuitively reason about statistical events they have never experienced before. Hence, they do not only possess a simple frequency detection mechanism, as suggested by the frequentist approach for human probabilistic reasoning

(Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995) and as has been found in various animal species (e.g. Behrend & Bitterman, 1961; Bullock & Bitterman, 1962; Graf et al., 1964; Longo, 1964; Luttbeg & Warner, 1999; Meyer, 1960; Real, 1991; Staddon, 1988; Wilson, 1960; Wodinsky et al. 1958). Moreover, great apes´ statistical abilities do not seem to be a product of heuristics or perceptual biases; they truly considered proportional information instead of absolute quantities (Eckert, Call, et al., 2018; Eckert, Rakoczy, et al., 2018; Rakoczy et al., 2014), and they did not follow the two most common heuristics described for humans: representativeness (Tversky and Kahneman, 1972) and availability (Tversky & Kahneman, 1974). More specifically, apes only expected the sample to resemble the population´s distribution when appropriate (when drawing was non-random, chimpanzees inferred that the sample instead should reflect the experimenter´s intentions; Eckert, Rakoczy, et al., 2018), contrasting predictions derived from the representative heuristic. Furthermore, they were able to spontaneously adjust their inferences to the current situation independent of their most recent experience, contrasting predictions derived from the availability heuristic. For example, when a biased experimenter suddenly drew blindly, chimpanzees intuitively ignored the events of the experience phase (where they learned about the experimenter´s biases) and chose based on the populations´ proportions from the first trial onwards (Eckert, Rakoczy, et al., 2018). Apes hence behaved like optimal reasoners: they were able to decide which of multiple cues of a scene were relevant in a particular situation and weighed their influence on the outcome of the current event. They adapted these evaluations dynamically: small changes in a developing situation (e.g. the presence of a screen as indication for blind drawing) appropriately changed the relative importance of other relevant cues (e.g. knowledge about experimenters´ biases). Our research also demonstrate that chimpanzees seem to have a basic grasp of randomness (Eckert, Rakoczy, et al., 2018). This finding is especially intriguing, since traditionally a full understanding of chance and randomness was called into question even for human adults (e.g. Gilovich et al., 1985; Kahneman & Tversky, 1972; Tversky & Kahneman, 1993).

One crucial indication of the present findings is that chimpanzees´ ability to flexibly consider knowledge from the social domain (Eckert, Rakoczy, et al., 2018) might be a first hint that apes´ statistical inference mechanism may be domain-general. If that was truly the case, great apes, just like human infants, possessed a powerful tool allowing them to rapidly acquire knowledge about their environment by drawing general conclusions from sparse data and to use these generalizations to predict the outcome of events in a variety of different contexts. This hypothesis is especially relevant in the light of more general debates about commonalities and differences in the cognitive architecture of humans and

nonhuman primates. Many prominent theories in comparative cognitive science postulate that one of the crucial differences between human and nonhuman cognition is based on the uniquely human forms of domain-general, flexible and systematic integration of different types of information (e.g. Carey, 2009; Penn, Holyoak, & Povinelli, 2008; Spelke, 2003; Tomasello & Call, 1997). The results obtained in the course of this dissertation cautiously suggest that great apes may actually be able to flexibly integrate different types of information to draw domain-general inferences, which would clearly contrast these theories. Future research will need to further explore this idea by investigating, e.g., whether great apes can integrate different (e.g. physical) information types in order to draw rational inferences.

Although the present dissertation highlights the flexibility of great apes´ intuitive statistics, at the same time, it falls short in unambiguously determining the generality of great apes´ abilities, in particular apes´ ability to reason from sample to population based on proportional information (Eckert et al., 2017). These limitations may rather be due to methodological short-comings than to restrictions in apes´ statistical abilities (see 6.1. and 7.1.1. for discussion). Nevertheless, in order to be able to draw any valid conclusions about a potential domain-general statistical inference mechanism, prospective research will have to further explore great apes´ capacity to draw inferences from sample to population in studies eliminating these extraneous task demands.

7.2.3. Implications for the debate on humans´ statistical reasoning (in-)abilities

A considerable part of the introduction of this thesis was devoted to research demonstrating the incapacity of human adults to draw rational statistical inferences (see sections 2.1. and 2.2.). The heuristics and biases program demonstrated that humans thoughtlessly rely on a small number of general-purpose heuristics leading to systematic errors and biases such as base-rate neglect, conjunction fallacy, recency effects, and an inability to account for sample size (Kahneman & Tversky, 1973, 1974; Tversky and Kahneman, 1971, 1972, 1983). The frequentist approach, on the other side, granted humans the ability to reason statistically, but only in very limited contexts, namely only as long as they can rely on frequencies of past experiences (e.g. Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). How does this fit the findings that pre-verbal infants possess an intuitive statistical inference mechanism? And, even more astonishingly, with the findings of the present dissertation demonstrating that even nonhuman great apes can intuitively reason from population to sample in quite sophisticated ways, without relying on either past experiences or heuristics? As Alison Gopnik (2014) phrased this apparent paradox: "Why are grown-ups often so stupid about probabilities when even babies and chimps can be so smart?".

As described above (see section 2.1.), the heuristics and biases program argued that probabilistic reasoning is a product of two distinct cognitive systems, in accordance with the widely held dual process account (Evans, 2003, 2008). According to this dual process account, the automatic, nonverbal and evolutionary ancient System 1 quickly responds to situations and easily leads to flawed and biased intuitions. In case they are not corrected by the more controlled and self-aware, verbal and explicit System 2 (for instance because of time pressure or cognitive load (e.g. Finucane et al, 2000; Gilbert, 1989; Bodenhausen, 1990)), humans´ judgments are biased and errors are committed (see e.g. Kahneman & Frederick, 2002 for a review). However, per definition, neither pre-verbal human infants nor nonhuman primates possess a System 2; consequently, their probability assessments should entirely be attributed to the unreliable and biased System 1. Nevertheless, their statistical inferences are not driven by simple general-purpose heuristics nor do they commit systematic errors as has been found in human adults, thereby giving counter evidence for the dual system theory proposed to account for this type of reasoning. Tversky and Kahneman (1971) argued that intuitions (which are the source of heuristics) should always be regarded "with proper suspicion" (p. 31). Does the evidence accumulated by developmental and comparative research suggest that intuitions can, in fact, be more rational than explicit and controlled judgments?

The apparently inconsistent findings from human adults on the one side and human infants and great apes on the other side, are probably, at least partly, caused by the different task formats usually applied for verbal and non-verbal creatures. The described developmental studies as well as our comparative studies involved, naturally, an experience-based format. Hence, infants and apes directly observed the situation they were required to make inferences about. In contrast, most of the previously described studies testing human adults used a descriptive format. More specifically, adults are usually confronted with text-based information involving explicit numerical information (e.g. percentages or fractions). It is possible that such description-based formats are masking human adults´ intuitive abilities in probabilistic reasoning tasks, whereas experience-based formats would facilitate them (see Schulze & Hertwig, unpublished, for a similar hypothesis). Eckert, Call, et al. (2018), for example, showed that human adults unambiguously succeed when confronted with an experience-based task comparable to that given to both great apes and human infants, supporting the hypothesis that task-format determines performance in statistical reasoning tasks. Moreover, developmental research has shown that young children regularly succeed in intuitive statistical tasks but fail as soon as there is a language component involved, e.g. when they are asked to verbally justify their judgments (e.g. Piaget & Inhelder, 1975; Teglas

et al., 2007). This suggests that the mechanism enabling intuitive statistics is completely independent of language, and thus might not be applied in tasks using a description-based format. In accordance with this hypothesis, there already is some evidence indicating that experience-based and description-based tasks can induce systematically different choices in adults as well as children (e.g. Barron & Erev, 2003; Camilleri & Newell, 2013; Hertwig & Erev, 2009). This "description-experience-gap" (Hertwig & Erev, 2009) is probably enhanced by the fact that text-based, symbolic probabilistic information (such as 30% or 1/3) require a true and exact understanding of fractions, a capacity that engages our symbolic number system. Hence, such kinds of inferences rely on a culturally and formally learned system (which is by definition not intuitive), in contrast to directly observed or experienced events, which apparently engage the evolutionary ancient analogue magnitude system instead. Therefore, if tested in the same, intuitive paradigms as infants and nonhuman primates, human adults would probably excel at least at similar, if not higher levels.

However, this does not seem to be the full story. Recent research presented young children as well as adults with variants of the "lawyer-engineer problem" (Kahneman & Tversky, 1973; see section 2.1.1.) to investigate the development of the representativeness heuristics (Gualtieri & Denison, 2018). Even though the paradigm was exactly the same for both children and adults (an experimental session was presented on a computer and narrated live by an experimenter, who then asked questions), results were different for the different age classes. More specifically, 4-year-olds produced responses that were closer to the base-rates, and hence more accurate than those of 5- to 6-year-old children and adults, who readily ignored base-rate information (6-year-olds and adults did so even more often than 5-year-olds). Thus, in comparison to older children and adults, young children seem to be less prone to neglect base-rates in favor of individuating information, suggesting that the representative heuristic only emerges later in human ontogeny. These findings are consistent with other studies demonstrating that younger children engage in less biased reasoning and pay more attention to current evidence than to prior assumptions and preconceptions (e.g. Lucas, Bridgers, Griffiths, & Gopnik, 2014). Hence, it is likely that general-purpose heuristics (and the accompanying biases and errors) may only develop with increasing experience and accumulation of factual knowledge. This, in turn, is consistent with the idea that, in fact, such heuristics are efficient and often effective tools, allowing humans to rapidly make judgments, especially in real world scenarios where parts of the relevant information are usually unknown (see, e.g. Gigerenzer & Gaissmaier, 2011 for a review). Following this hypothesis, it is unlikely that either infants or great apes engage in

general-purpose-heuristics such as representativeness. It will be of great interest for future research to address this hypothesis.

Taken together, the apparent discrepancy between human adult statistical abilities on the one side, and those of human infants and great apes on the other side are probably due to two (perhaps interacting) main factors: First, the description-based task format mostly applied in studies testing adults probably engages the symbolic number system instead of the intuitive analogue magnitude system which is applied by infants and great apes. Second, humans seem to generate mental-shortcuts as time-saving alternatives to statistical computations over the course of development. While such heuristics are usually efficient and sufficient in real-world scenarios, they can lead to the systematic biases and errors frequently observed in specific task designs.

In conclusion, recent developmental and comparative research, including the work of the present dissertation, suggests that humans possess an intuitive, evolutionary ancient non-symbolic statistical reasoning device, allowing us to draw statistical inferences from population to sample and vice versa. This system seems to be present from birth, and it is shared with our closest living relatives, the nonhuman great apes. Its signatures follow Weber´s law - hence, its probability estimates are not exact, but only approximate. From an evolutionary perspective, such an approximate statistical inference mechanism is probably sufficient for survival: For example, it is enough to have a rough intuition which of several foraging sites is more likely to have a sufficient amount of food and less likely to be invaded by predators - it is not necessary to know exactly how much more or less likely. Our modern society, however, needs to solve statistical problems in an "as close to exact as possible" way. For instance, coming back to the example mentioned in the introduction, before a new migraine treatment is being approved, it has to be investigated at exactly which rate it alleviates the symptoms and how likely potential side effects will occur. Similarly, in order to be able to draw any conclusion from the empirical data presented in this dissertation, I had to analyze them in a normative way, allowing us to determine exactly how likely it is that they were generated by a random process. Such higher-level statistical inferences are not intuitive at all, they require extensive training and would not be possible without our symbolic number system. Hence, much like "human´s explicit knowledge of symbolic number and geometry may be our unique, abstract cognitive ability (…), but it clearly depends, in part, on a system that is widely shared by other animals" (Vallortigara 2015, p. 59), also formal (i.e. symbolic) statistical reasoning may be our unique, abstract cognitive ability, but it is most likely, at least in part, based on an intuitive system that is shared with our closest living relatives.

7.3. Outlook

The current dissertation demonstrated that chimpanzees and other nonhuman great apes possess a cognitive mechanism for intuitive statistics comparable to that of human infants, which enables them to reason from population to sample and, to a certain extent, vice versa. These findings present important steps towards a better understanding of the evolutionary origins of intuitive statistics. At the same time, these new insights leave open and raise exciting new questions about the cognitive underpinnings of intuitive statistics that need to be systematically addressed. Some of these questions tackle the more detailed nature of great apes´ statistical abilities, others the phylogenetic distribution of such abilities, and yet others the cognitive structure and development of intuitive statistics in general.

7.3.1. Open questions concerning the nature of great apes´ statistical abilities

We are only beginning to understand the statistical abilities of nonhuman primates and, accordingly, there is a wide variety of unanswered research questions to be addressed, some of which I will list in this section. One important open question which could not be fully answered within this dissertation concerns great apes´ reasoning capacities from sample to population. True statistical inferences involve a genuine understanding of the relation between populations, sampling processes and resulting samples (see Figure 2). To date, it remains unclear whether great apes can truly reason from sample to population based on proportional information (Eckert et al., 2017), even though some of our work points in that direction (Eckert, Rakoczy, et al., 2018). In fact, even for human infants, the existing studies testing for such abilities (Denison et al. 2012; Xu & Garcia, 2008) leave open some important questions: Are infants able to reason from sample to population in an active-choice paradigm? And do they succeed even when absolute and relative frequencies are disentangled? The ability to reason statistically from sample to population is particularly important, since it allows an individual to draw general rules from limited observations, and hence depicts a likely candidate mechanism for inductive learning. Prospective research, therefore, should develop a more comparable paradigm for great apes and infants, carefully disentangling absolute and relative frequencies, and reducing extraneous task demands such as working memory load, inhibitory control demands and small magnitudes of differences between ratios, to more comprehensively explore both great apes´ and human infants´ abilities to reason statistically from sample to population.

A second set of questions raised by the previous findings is whether great apes, like human infants, can integrate different kinds of information into their statistical inferences. Eckert, Rakoczy, et al.

(2018) showed that apes do have a random sampling assumption and that they recognize when this random sampling assumption is violated by experimenters who draw samples in an intentional way. These findings suggest that apes, just like human infants, flexibly combine knowledge from different cognitive domains with statistical input. At the same time, they raise a number of questions: First, how flexible and general are these capacities? Can apes use information other than social context information to judge whether the random sampling assumption is fulfilled? Developmental research has shown that infants can use a variety of different types of information when making inferences about the outcome of events (e.g. about different physical constraints (e.g. Denison & Xu, 2010; Teglas et al., 2007), or spatiotemporal information (Teglas et al., 2011)). For apes, so far only one study (Eckert et al., unpublished) has investigated whether apes can combine statistical and physical information (a physical barrier separating two types of food). Results showed that apes did understand that the physical constraint influenced the outcome of the sampling event. However, when physical and proportional information were in conflict, apes had difficulties understanding that the physical constraint determined the outcome regardless of the proportional composition of the population (e.g. that a population consisting of mainly peanuts was impossible to yield a peanut sample when all of them were stuck behind a barrier). It is possible, though, that these findings were due to the rather complicated task setup rather than reflecting true limitations in chimpanzees´ abilities. It, therefore, remains an open question whether great apes can use information about physical circumstances in combination with distributional information to predict the outcome of events in a more straightforward task setup. Second, are chimpanzees able to update statistical inferences in light of new evidence to compute a posterior probability? Girotto and Gonzales (2008) showed that from age 5, children´s decisions under uncertainty are sensitive to posterior information: Children were confronted with a population of chips; half of them were squares, the others were round. While all round chips were blue, most, but not all, squares were red. When the experimenter drew one item and asked which color it had, children correctly stated that it was probably blue. When, however, the experimenter informed them that he had drawn a square, children revised their prediction and answered that the sample was probably red. Hence, children were able to integrate prior probability with additional information to compute a posterior probability. Pre-literate Mayan adults succeeded in a similar version of this task, despite their lack of any sort of formal education (Fontanari et al., 2014). To my knowledge, there are no studies testing whether pre- or non-verbal individuals can update their statistical inferences in light of new evidence. It will be an exciting avenue of research to investigate whether apes and human infants are capable of such advanced intuitive statistical abilities.

A third set of questions concerns great apes´ ability to apply statistical capacities in more natural situations. The current dissertation demonstrated that apes have sophisticated intuitive statistical reasoning abilities, seemingly on a par with those of human infants. However, to date all studies utilized a very similar paradigm that required subjects to use proportional information within a population of food items to predict the outcome of a human´s drawing process (or vice versa). It remains an open question whether and how apes actually use this statistical tool kit outside the much used and rather unnatural "bucket paradigm". Chimpanzees live in complex social groups characterized by both agonistic and mutualistic relationships: While chimpanzees are generally a competitive species (e.g. Hare, 2001; Muller & Mitani, 2005), they are also able to successfully cooperate with their group mates in order to achieve mutual gains (see, e.g. Mitani, 2009; Schmelz & Call, 2016 for reviews). Being able to infer and predict the behavior of other individuals can therefore be highly beneficial, both in order to outperform others, and in order to affiliate with the most reliable partners in cooperative events. A variety of studies have demonstrated that chimpanzees have flexible social cognitive abilities and share with humans the capacity to attribute (at least some) mental states to others (for a review see, e.g. Call & Tomasello, 2008). Furthermore, chimpanzees were shown to use these "mind reading" abilities to judge the behavior of others, both in cooperative (e.g. Yamamoto, Humle, & Tanaka, 2012) and in competitive situations (e.g. Kaminski, Call, & Tomasello, 2008; Schmelz, Call, & Tomasello, 2011, 2013). I hypothesize that, to make these judgments, apes apply their statistical inference mechanism: using frequency information as input and integrating it with other types of information, statistical reasoning may be a powerful tool to compute event probabilities and to draw domain-general inferences. In the context of competition, chimpanzees may use proportional information to infer habits and preferences of others and use this information to avoid conflict. Support for this idea comes from the findings of this dissertation: Eckert, Call, et al. (2018) showed that chimpanzees inferred humans´ choice biases from a violation of statistical likelihoods in the standard bucket paradigm. It thus seems plausible that apes would also use such statistically acquired knowledge in competitive situations. In the context of cooperation and collaboration, apes may use proportional information to assess the competence or reliability of conspecifics. Melis, Hare, and Tomasello (2006) found that chimpanzees recruited the more effective of two partners for a collaborative task on the basis of their prior experience with each of them. It will be an intriguing question to investigate whether apes use the proportion of previous successes/failures rather than the absolute number of successes to make their decision. Developmental research has shown that human children may use intuitive statistical reasoning for social learning (Pasquini, Corriveau, Koenig, & Harris, 2007): Children chose who to trust based on previous accuracy of informants; while initially this selective trust seems to

work in all-or-nothing-fashion, children from age 4 tracked the relative frequency of errors and used this proportion of failures to successes to predict future accuracy of informants. Prospective research will need to explore whether chimpanzees and other great apes perform on the level of human infants and hence if their reliance on conspecifics also works in all-or-nothing-fashion, or whether they, like older children, track the relative frequency of errors in order to predict future success.

Another interesting avenue for future research on great apes´ intuitive statistics addresses the relationship between randomness and fairness. For humans, randomness is a very useful concept and an indispensable basis for fair play: Randomly shuffling cards, for example, ensures equal chances for all players to get a good hand. This procedure is indisputably fair, even if it may result in unevenly distributed winning cards. Eckert, Rakoczy, et al. (2018) showed that chimpanzees can discriminate random from non-random drawing, giving a first hint that great apes may, too, have an understanding of randomness. Do great apes also consider randomness, in contrast to intention, as basis of "fairness"? This question is particularly exciting in light of the debate on whether or not nonhuman primates have a sense of fairness regarding the distribution of food items among conspecifics (e.g. Bräuer, Call, & Tomasello, 2006, 2009; Brosnan & Waal, 2003, 2014; Engelmann, Clift, Herrmann, & Tomasello, 2017; Wynne, 2004). One way to further inform this debate is to investigate whether apes react differently in response to uneven reward distribution depending on whether a human drew and distributed the rewards intentionally or randomly.

7.3.2. Open questions concerning the phylogenetic distribution of statistical abilities

The current dissertation only scratches the surface of explaining the evolutionary roots of intuitive statistics. While the research so far unequivocally demonstrates that statistical reasoning abilities emerged before the separation of humans and chimpanzees (and most likely other nonhuman great apes), it remains unclear whether similar capacities exist outside the great ape lineage. As described above, so far there is no unambiguous evidence for statistical reasoning abilities in monkeys (Placi et al., submitted; Tecwyn, et al., 2016). However, considering that the analogue magnitude system, which seems to be the foundation of intuitive statistics (Eckert, Call, et al.; 2018), is present in a wide range of species and taking into account that these abilities most likely bring clear fitness benefits, it is reasonable to assume that statistical abilities may have emerged several times independently during animal evolution.

To understand the evolutionary roots of intuitive statistics it is therefore essential to consider the possibility of convergent evolution and test cognitively flexible species outside the primate lineage.

Finding analogue traits in distantly related species will give us important insights in both the ecological pressures and the cognitive prerequisites that facilitated and enabled the evolution of intuitive statistics. As described above (see section 7.2.2.), two recent studies found evidence for sensitivity to probabilistic reward patterns based on proportional information in pigeons and a grey parrot (Clements et al., 2018; Roberts et al., 2018). These findings strengthen the hypothesis that intuitive statistical reasoning abilities may have evolved independently inside *and* outside the primate order. To investigate this hypothesis more systematically and comprehensively, a larger number of individuals should be tested in an intuitive task which allows to directly compare with primates´ performance.

7.3.3. Open questions concerning the cognitive structure and development of intuitive statistical abilities

Eckert, Call, et al. (2018) demonstrated that the analogue magnitude system most likely underlies intuitive statistical reasoning abilities both in chimpanzees and human adults. This finding raises a couple of questions concerning the exact nature of this mechanism:

First, does this mechanism compute probabilities over individual objects, i.e. discrete quantities, or rather over continuous magnitudes such as surface area or volume? In the studies conducted in the course of this dissertation, subjects could have reasoned about proportions of discrete quantities (e.g. a population of 40 peanuts vs. 10 carrot pieces depicts an 80% chance of drawing a peanut as a random sample) or about proportions of continuous quantities (e.g. the total volume of a bucket is filled 80% with peanuts and 20% with carrots, so chances of drawing a peanut are likewise 80%). Similarly, infants in the previously described studies (e.g. Denison & Xu, 2010a, 2014) might have reasoned about, e.g. the relative frequencies of red and white Ping-Pong balls, or about the proportion of the colors red and white. Theoretically, both options are conceivable: As mentioned above, many studies investigating quantitative abilities in nonhuman primates (and likewise in infants) controlled for various continuous dimensions such as surface area or duration and found that great apes and monkeys indeed do possess the ability to represent discrete numerical information (e.g., Beran, 2007; Cantlon & Brannon, 2007a; Thomas et al., 1980). Similarly, they are also able to reason about continuous quantities, such as amounts of liquids (see, e.g. Beran, 2010; Call & Rochat, 1996; Muncer, 1983; Suda & Call, 2004, 2005). Both types of representations are most likely mediated by an analogue magnitude system (see, e.g. Cantlon et al., 2009 and Lourenco, 2015 for reviews), signatures of which were also found in tasks requiring intuitive statistical

inferences (Eckert, Call, et al., 2018; for details on the debate on whether there is a separate approximate number system which is supporting only discrete numerical judgments, see section 3.3.1.2.). For human children, research found some evidence that proportional reasoning seems to be easier when dealing with continuous compared to discrete quantities (Boyer, Levine, & Huttenlocher, 2008; Jeong, Levine, & Huttenlocher, 2007; Spinillo & Bryant, 1999), giving a first hint that intuitive statistics might primarily be computed over continuous magnitudes. Future studies should test both human infants and great apes in a statistical inference task disentangling continuous and discrete quantities to further inform this debate.

A second set of questions regarding the cognitive structure of intuitive statistics concerns its developmental trajectory as well as its relation to formal (i.e. symbolic) statistical reasoning. For humans, research already shed some light on the developmental pathway of numerical abilities regarding absolute set sizes and found some interesting patterns. First, there seem to be large individual differences in the accuracy of the analogue magnitude system (Halberda, Mazzocco, & Feigenson, 2008), which appear to be consistent over development (Starr, Libertus, & Brannon, 2013). In fact, early inter-individual differences in accuracy even seem to be predictive of later explicit mathematical achievement (Feigenson et al., 2013; Libertus, Feigenson, & Halberda, 2013b; Mazzocco, Feigenson, & Halberda, 2011; but see Gilmore et al., 2011; Libertus et al., 2013a). Moreover, the system´s accuracy seems to increase during childhood, with lowest precision levels in infancy which improve over the course of development until adulthood (Halberda & Feigenson, 2008; Pica et al., 2004; Xu & Spelke, 2000). These findings raise the question whether the development of statistical abilities follows similar patterns. More specifically, do we find better accuracy in statistical reasoning tasks in older compared to younger individuals? And are early inter-individual differences in accuracy predictive of later formal statistical reasoning? As described earlier, there already is some evidence contrasting these predictions. One study (Girotto et al., 2016), for example, found that 3- to 4-year-old children failed in intuitive statistical tasks, in which pre-verbal infants repeatedly succeeded (Denison & Xu, 2010a, 2014), even when verbal demands were reduced. This may be a first hint that, in contrast to abilities dealing with absolute quantities, intuitive statistics develop in a non-linear way. Future research will have to examine whether the older children's failure in this study was truly due to cognitive limitations, or rather due to performance limitations caused by extraneous task demands. Another hint pointing towards a nonlinear development of intuitive statistics comes from a study demonstrating that 4-year-old children formed more rational inferences (considering base-rate information), compared to older children and adults (Gualtieri & Denison, 2018). To my knowledge, there are no studies investigating the developmental trajectory of any quantitative ability in nonhuman great

apes. Moreover, all apes tested in intuitive statistical tasks, so far, were adults or adolescents. In order to shed more light on the developmental trajectory of intuitive statistics, ideally, one would need to conduct longitudinal studies with both humans and great apes, using the same active-choice paradigm at certain time points throughout life stages starting in infancy with continuous tests until adulthood.

In conclusion, the present dissertation depicts an important first step in investigating the evolutionary roots of intuitive statistics. It demonstrates that exploring such capacities in nonhuman great apes has enormous potential for gaining insights both into the origins of human statistical reasoning, as well as into the cognitive architecture of our closest living relatives. I am sanguine that this work will serve as a stepping stone to stimulate more research in this promising new field of comparative cognition.

8. References

Abramson, J. Z., Hernández-Lloreda, V., Call, J., & Colmenares, F. (2013). Relative quantity judgments in the beluga whale (*Delphinapterus leucas*) and the bottlenose dolphin (*Tursiops truncatus*). *Behavioural Processes*, *96*, 11–19. https://doi.org/10.1016/j.beproc.2013.02.006

Acredolo, C., O'Connor, J., Banks, L., & Horobin, K. (1989). Children's Ability to Make Probability Estimates: Skills Revealed Through Application of Anderson's Functional Measurement Methodology. *Child Development*, *60*(4), 933–945. https://doi.org/10.2307/1131034

Adachi, J., & Hasegawa, M. (1995). Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: Heterogeneity among amino acid sites. *Journal of Molecular Evolution*, *40*(6), 622–628. https://doi.org/10.1007/BF00160510

Addessi, E., Crescimbene, L., & Visalberghi, E. (2008). Food and token quantity discrimination in capuchin monkeys (*Cebus apella*). *Animal Cognition*, *11*(2), 275–282. https://doi.org/10.1007/s10071-007-0111-6

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2008). Do fish count? Spontaneous discrimination of quantity in female mosquitofish. *Animal Cognition*, *11*(3), 495–503. https://doi.org/10.1007/s10071-008-0140-9

Agrillo, C., Piffer, L., & Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition*, *119*(2), 281–287. https://doi.org/10.1016/j.cognition.2010.10.022

Aguiar, A., & Baillargeon, R. (1999). 2.5-Month-Old Infants' Reasoning about When Objects Should and Should Not Be Occluded. *Cognitive Psychology*, *39*(2), 116–157. https://doi.org/10.1006/cogp.1999.0717

Ahmed, A., & Ruffman, T. (1998). Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a nonsearch task? *Developmental Psychology*, *34*(3), 441–453. https://doi.org/10.1037//0012-1649.34.3.441

Ain, S. A., Giret, N., Grand, M., Kreutzer, M., & Bovet, D. (2009). The discrimination of discrete and continuous amounts in African grey parrots (*Psittacus erithacus*). *Animal Cognition*, *12*(1), 145–154. https://doi.org/10.1007/s10071-008-0178-8

Anderson, U. S., Stoinski, T. S., Bloomsmith, M. A., & Maple, T. L. (2007). Relative numerousness judgment and summation in young, middle-aged, and older adult orangutans (*Pongo pygmaeus abelii* and *Pongo pygmaeus pygmaeus*). *Journal of Comparative Psychology (Washington, D.C.: 1983)*, *121*(1), 1–11. https://doi.org/10.1037/0735-7036.121.1.1

Anderson, U. S., Stoinski, T. S., Bloomsmith, M. A., Marr, M. J., Smith, A. D., & Maple, T. L. (2005). Relative numerousness judgment and summation in young and old Western lowland gorillas. *Journal of Comparative Psychology (Washington, D.C.: 1983)*, *119*(3), 285–295. https://doi.org/10.1037/0735-7036.119.3.285

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, *32*(8), 1369–1378. https://doi.org/10.3758/BF03206327

Baillargeon, R. (2008). Innate Ideas Revisited: For a Principle of Persistence in Infants' Physical Reasoning. *Perspectives on Psychological Science*, *3*(1), 2–13. https://doi.org/10.1111/j.1745-6916.2008.00056.x

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191–208. https://doi.org/10.1016/0010-0277(85)90008-3

Balakrishnan, J. D., & Ashby, F. G. (1991). Is subitizing a unique numerical ability? *Perception & Psychophysics*, *50*(6), 555–564. https://doi.org/10.3758/BF03207540

Balakrishnanl, J. D., & Ashby, F. G. (1992). Subitizing: Magical numbers or mere superstition? *Psychological Research*, *54*(2), 80–90. https://doi.org/10.1007/BF00937136

Balci, F., Freestone, D., & Gallistel, C. R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Sciences*, pnas.0812709106. https://doi.org/10.1073/pnas.0812709106

Barnard, A. M., Hughes, K. D., Gerhardt, R. R., Divincenti, L., Jr., Bovee, J. M., & Cantlon, J. F. (2013). Inherently Analog Quantity Representations in Olive Baboons (*Papio anubis*). *Frontiers in Psychology*, *4*, 253. https://doi.org/10.3389/fpsyg.2013.00253

Barner, D., Wood, J., Hauser, M., & Carey, S. (2008). Evidence for a non-linguistic distinction between singular and plural sets in rhesus monkeys. *Cognition*, *107*(2), 603–622. https://doi.org/10.1016/j.cognition.2007.11.010

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.

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201–221. https://doi.org/10.1016/S0010-0277(02)00178-6

Behrend, E. R., & Bitterman, M. E. (1961). Probability-Matching in the Fish. *The American Journal of Psychology*, *74*(4), 542–551. https://doi.org/10.2307/1419664

Benson-Amram, S., Heinen, V. K., Dryer, S. L., & Holekamp, K. E. (2011). Numerical assessment and individual call discrimination by wild spotted hyaenas, *Crocuta crocuta*. *Animal Behaviour*, *82*(4), 743–752. https://doi.org/10.1016/j.anbehav.2011.07.004

Beran, M. J. (2001). Summation and numerousness judgments of sequentially presented sets of items by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *115*(2), 181–191. https://doi.org/10.1037/0735-7036.115.2.181

Beran, M. J. (2004). Chimpanzees (*Pan troglodytes*) respond to nonvisible sets after one-by-one addition and removal of items. *Journal of comparative psychology (Washington, DC: 1983)*, *118*(1), 25.

Beran, M. J. (2007). Rhesus Monkeys (*Macaca mulatta*) Enumerate Large and Small Sequentially Presented Sets of Items Using Analog Numerical Representations. *Journal of Experimental Psychology. Animal Behavior Processes*, *33*(1), 42–54. https://doi.org/10.1037/0097-7403.33.1.42

Beran, M. J. (2010). Chimpanzees (*Pan troglodytes*) accurately compare poured liquid quantities. *Animal Cognition*, *13*(4), 641–649. https://doi.org/10.1007/s10071-010-0314-0

Beran, M. J. (2012). Quantity Judgments of Auditory and Visual Stimuli by Chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology. Animal Behavior Processes*, *38*(1), 23–29. https://doi.org/10.1037/a0024965

Beran, M. J. (2017). Quantitative Cognition. In J. Call (Ed.), *APA Handbook of Comparative Psychology.* Vol 2: Perception, learning, and cognition (pp. 535-577). Washington, DC: American Psychological Association

Beran, M. J., Evans, T. A., Leighty, K. A., Harris, E. H., & Rice, D. (2008). Summation and quantity judgments of sequentially presented sets by capuchin monkeys (*Cebus apella*). *American Journal of Primatology*, *70*(2), 191–194. https://doi.org/10.1002/ajp.20474

Biernaskie, J. M., Walker, S. C., & Gegear, R. J. (2009). Bumblebees Learn to Forage like Bayesians. *The American Naturalist*, *174*(3), 413–423. https://doi.org/10.1086/603629

Bodenhausen, G. V. (1990). Stereotypes as Judgmental Heuristics: Evidence of Circadian Variations in Discrimination. *Psychological Science*, *1*(5), 319–322. https://doi.org/10.1111/j.1467-9280.1990.tb00226.x

Bonanni, R., Natoli, E., Cafazzo, S., & Valsecchi, P. (2011). Free-ranging dogs assess the quantity of opponents in intergroup conflicts. *Animal Cognition*, *14*(1), 103–115. https://doi.org/10.1007/s10071-010-0348-3

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Bonawitz, E., Fischer, A., & Schulz, L. (2011). Teaching three-and-a-half-year-olds to reason about ambiguous evidence. *MIT Web Domain*. Retrieved from http://dspace.mit.edu/handle/1721.1/60991

Bonini, N., Tentori, K., & Osherson, D. (2004). A Different Conjunction Fallacy. *Mind & Language*, *19*(2), 199–210. https://doi.org/10.1111/j.1468-0017.2004.00254.x

Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology*, *114*(3), 375–388. https://doi.org/10.1016/j.jecp.2012.09.015

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, *44*(5), 1478–1490. https://doi.org/10.1037/a0013110

Boysen, S. T., & Berntson, G. G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology*, *103*(1), 23–31.

Boysen, S. T., & Berntson, G. G. (1995). Responses to Quantity - Perceptual Versus Cognitive Mechanisms in Chimpanzees (*Pan-Troglodytes*). *Journal of Experimental Psychology-Animal Behavior Processes*, *21*(1), 82–86. https://doi.org/10.1037/0097-7403.21.1.82

Boysen, S. T., Berntson, G. G., & Hannah, M. B. (1996). Quantity-based interference and symbolic representations in chimpanzees (Pan troglodytes. *Journal of Experimental Psychology: Animal Behavior Processes*, 76–86.

Bradbeer, J. W. (1988). Seed dormancy and germination. *Seed Dormancy and Germination.* Retrieved from https://www-cabdirect-org.offsitelib.eva.mpg.de/cabdirect/abstract/19880717835

Brannon, E. M., Lutz, D., & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. *Developmental Science*, *9*(6), F59–F64. https://doi.org/10.1111/j.1467-7687.2006.00530.x

Brannon, E. M., Suanda, S., & Libertus, K. (2007). Temporal Discrimination Increases in Precision over Development and Parallels the Development of Numerosity Discrimination. *Developmental Science*, *10*(6), 770-777.

Brannon, E. M., & Terrace, H. S. (2000). Representation of the numerosities 1-9 by rhesus macaques (Macaca mulatta). *Journal of Experimental Psychology. Animal Behavior Processes*, *26*(1), 31–49.

Brannon, E. M., Wusthoff, C. J., Gallistel, C. R., & Gibbon, J. (2001). Numerical Subtraction in the Pigeon: Evidence for a Linear Subjective Number Scale, *Psychological Science*, *12*(3), 238–243. https://doi.org/10.1111/1467-9280.00342

Bräuer, J., Call, J., & Tomasello, M. (2006). Are apes really inequity averse? *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1605), 3123–3128. https://doi.org/10.1098/rspb.2006.3693

Bräuer, J., Call, J., & Tomasello, M. (2009). Are apes inequity averse? New data on the token-exchange paradigm. *American Journal of Primatology*, *71*(2), 175–181. https://doi.org/10.1002/ajp.20639

Brosnan, S. F., & Waal, F. B. M. de. (2003). Monkeys reject unequal pay. *Nature*, *425*(6955), 297–299. https://doi.org/10.1038/nature01963

Brosnan, S. F., & Waal, F. B. M. de. (2014). Evolution of responses to (un)fairness. *Science*, *346*(6207), 1251776. https://doi.org/10.1126/science.1251776

Buckingham, J. N., Wong, B. B. M., & Rosenthal, G. G. (2007). Shoaling decisions in female swordtails: how do fish gauge group size? *Behaviour*, *144*, 1333–1346. https://doi.org/10.1163/156853907782418196

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition*, *118*(1), 32–44. https://doi.org/10.1016/j.cognition.2010.09.005

Bullock, D. H., & Bitterman, M. E. (1962). Probability-Matching in the Pigeon. *The American Journal of Psychology*, *75*(4), 634–639. https://doi.org/10.2307/1420288

Butterworth, B. (1999). What counts: how every brain is hardwired for math. *New York, USA: The Free Press*. Retrieved from http://discovery.ucl.ac.uk/4372/

Call, J. (2000). Estimating and operating on discrete quantities in orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology*, *114*(2), 136–147. https://doi.org/10.1037//0735-7036.114.2.136

Call, J., & Rochat, P. (1996). Liquid conservation in orangutans (*Pongo pygmaeus*) and humans (*Homo sapiens*): individual differences and perceptual strategies. *Journal of Comparative Psychology (Washington, D.C.: 1983)*, *110*(3), 219–232.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends Cognitive Science*, *12*(5), 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Camilleri, A. R., & Newell, B. R. (2013). Chapter 4 - Mind the gap? Description, experience, and the continuum of uncertainty in risky choice. In V. S. C. Pammi & N. Srinivasan (Eds.), *Progress in Brain Research* (Vol. 202, pp. 55–71). Elsevier. https://doi.org/10.1016/B978-0-444-62604-2.00004-6

Cantlon, J. F. (2015). Chapter 9 - Analog Origins of Numerical Concepts. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Mathematical Cognition and Learning* (Vol. 1, pp. 225–251). Elsevier. https://doi.org/10.1016/B978-0-12-420133-0.00009-0

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychol Sci*, *17*(5), 401–406. https://doi.org/10.1111/j.1467-9280.2006.01719.x

Cantlon, J. F., & Brannon, E. M. (2007a). How much does number matter to a monkey (Macaca mulatta)? *Journal of Experimental Psychology. Animal Behavior Processes*, *33*(1), 32–41. https://doi.org/10.1037/0097-7403.33.1.32

Cantlon, J. F., & Brannon, E. M. (2007b). Basic math in monkeys and college students. *PLoS Biol*, *5*(12), e328. https://doi.org/10.1371/journal.pbio.0050328

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional Imaging of Numerical Processing in Adults and 4-y-Old Children. *PLOS Biology*, *4*(5), e125. https://doi.org/10.1371/journal.pbio.0040125

Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on" Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures". *science*, *323*(5910), 38-38.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83–91. https://doi.org/10.1016/j.tics.2008.11.007

Cantlon, J. F., Safford, K. E., & Brannon, E. M. (2010). Spontaneous analog number representations in 3-year-old children. *Developmental Science*, *13*(2), 289–297. https://doi.org/10.1111/j.1467-7687.2009.00887.x

Carey, S. (2009). The origin of concepts. New York, NY: *Oxford University Press*; US

Carey, S., & Spelke, E. (1996). Science and Core Knowledge. *Philosophy of Science*, *63*(4), 515–533. https://doi.org/10.1086/289971

Carnap, R. (1945). The Two Concepts of Probability: The Problem of Probability. *Philosophy and Phenomenological Research*, *5*(4), 513–532. https://doi.org/10.2307/2102817

Cesana-Arlotti, N., Téglás, E., & Bonatti, L. L. (2012). Chapter One - The Probable and the Possible at 12 Months: Intuitive Reasoning about the Uncertain Future. In Fei Xu & T. Kushnir (Eds.), *Advances in Child Development and Behavior* (Vol. 43, pp. 1–25). JAI. https://doi.org/10.1016/B978-0-12-397919-3.00001-0

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, *74*(3), 271–280.

Charness, G., Karni, E., & Levin, D. (2010). On the conjunction fallacy in probability judgment: New experimental evidence regarding Linda. *Games and Economic Behavior*, *68*(2), 551–556. https://doi.org/10.1016/j.geb.2009.09.003

Chase, R. H. V. M. (1998). Many Reasons or Just One: How Response Mode Affects Reasoning in the Conjunction Problem. *Thinking & Reasoning*, *4*(4), 319–352. https://doi.org/10.1080/135467898394102

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, *2*(6), 206–214. https://doi.org/10.1016/S1364-6613(98)01179-6

Chater, N., & Oaksford, M. (2008). The Probabilistic Mind: Prospects for Bayesian Cognitive Science. *Oxford University Press.*

Chiesi, F., & Primi, C. (2009). Recency effects in primary-age children and college students. *IEJME-Mathematics Education*.

Choo, H., & Franconeri, S. L. (2014). Enumeration of small collections violates Weber's law. *Psychonomic Bulletin & Review*, *21*(1), 93–99. https://doi.org/10.3758/s13423-013-0474-4

Church, R. M. (1984). The numerical attribute of stimuli. *Animal Cognition*. Retrieved from
https://ci.nii.ac.jp/naid/10012745189/

Clearfield, M. W., & Mix, K. S. (1999). Number Versus Contour Length in Infants' Discrimination of Small
Visual Sets. *Psychological Science*, *10*(5), 408–411. https://doi.org/10.1111/1467-9280.00177

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

Collins, E. J., McNamara, J. M., & Ramsey, D. M. (2006). Learning rules for optimal selection in a varying
environment: mate choice revisited. *Behavioral Ecology*, *17*(5), 799–809.
https://doi.org/10.1093/beheco/arl008

Cordes, S., & Brannon, E. M. (2009). Crossing the Divide: Infants Discriminate Small from Large
Numerosities. *Developmental Psychology*, *45*(6), 1583–1594. https://doi.org/10.1037/a0015666

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

Costello, F. J. (2009). How probability theory explains the conjunction fallacy. *Journal of Behavioral
Decision Making*, *22*(3), 213–234. https://doi.org/10.1002/bdm.618

Cresswell, W. (1994). Flocking is an effective anti-predation strategy in redshanks, *Tringa totanus*.
Animal Behaviour, *47*(2), 433–442. https://doi.org/10.1006/anbe.1994.1057

Dacke, M., & Srinivasan, M. V. (2008). Evidence for counting in insects. *Animal Cognition*, *11*(4), 683–
689. https://doi.org/10.1007/s10071-008-0159-y

Dadda, M., Piffer, L., Agrillo, C., & Bisazza, A. (2009). Spontaneous number representation in
mosquitofish. *Cognition*, *112*(2), 343–348. https://doi.org/10.1016/j.cognition.2009.05.009

Davidson, D. (1995). The Representativeness Heuristic and the Conjunction Fallacy Effect in Children's
Decision Making. *Merrill-Palmer Quarterly*, *41*(3), 328–346.

Davidson, D., & Hirtle, S. C. (1990). Effects of Nondiscrepant and Discrepant Information on the Use of
Base Rates. *The American Journal of Psychology*, *103*(3), 343–357.
https://doi.org/10.2307/1423214

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time,
and number in neonates. *Proceedings of the National Academy of Sciences*, *111*(13), 4809–4813.
https://doi.org/10.1073/pnas.1323628111

de Hevia, M. D., & Spelke, E. S. (2010). Number-Space Mapping in Human Infants. *Psychological Science*,
21(5), 653–660. https://doi.org/10.1177/0956797610366091

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (1997). The number sense: How mathematical knowledge is embedded in our brains. *Oxford University Press.*

Dehaene, S. (2003). The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147. https://doi.org/10.1016/S1364-6613(03)00055-X

Dehaene, S, & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(5), 958–975.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *science*, *320*(5880), 1217-1220.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology,* 20, 487-506.

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 naive 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). 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. (2010b). 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. (2012). Probabilistic inference in human infants. *Advances in Child Development and Behavior*, *43*, 27–58.

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

Descartes, R., & Lafleur, L. J. (1960). Meditations on first philosophy. Indianapolis: Bobbs-Merrill.

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00068

Diester, I., & Nieder, A. (2007). Semantic Associations between Signs and Numerical Categories in the Prefrontal Cortex. *PLOS Biology*, *5*(11), e294. https://doi.org/10.1371/journal.pbio.0050294

Dooley, G. B., & Gill, T. V. (1977). chapter 13 - Acquisition and Use of Mathematical Skills by a Linguistic Chimpanzee. In Duane M. Rumbaugh (Ed.), *Language Learning by a Chimpanzee* (pp. 247–260). Academic Press. https://doi.org/10.1016/B978-0-12-601850-9.50020-8

Droit-Volet, S., Clément, A., & Fayol, M. (2008). Time, number and length: Similarities and differences in discrimination in adults and children. *The Quarterly Journal of Experimental Psychology*, *61*(12), 1827–1846. https://doi.org/10.1080/17470210701743643

Drucker, C. B., Rossa, M. A., & Brannon, E. M. (2016). Comparison of discrete ratios by rhesus macaques (*Macaca mulatta*). *Animal Cognition*, *19*(1), 75–89. https://doi.org/10.1007/s10071-015-0914-9

Dulany, D. E., & Hilton, D. J. (1991). Conversational Implicature, Conscious Representation, and the Conjunction Fallacy. *Social Cognition*, *9*(1), 85–110. https://doi.org/10.1521/soco.1991.9.1.85

Ebel, S. J., & Call, J. (2018). The interplay of prior experience and motivation in great ape problem-solving (*Gorilla gorilla, Pan paniscus, Pan troglodytes, and Pongo abelii*). *Journal of Comparative Psychology*. https://doi.org/10.1037/com0000117

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. doi: 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). doi: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. doi: 10.1016/j.cub.2018.04.077

Elman, J., Bates, E., Johnson, M., Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press.

Em, B., S, S., & K, L. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Developmental Science*, *10*(6), 770–777.

Emmerton, J. (1998). Numerosity differences and effects of stimulus density on pigeons' discrimination performance. *Animal Learning & Behavior*, *26*(3), 243–256. https://doi.org/10.3758/BF03199218

Engelmann, J. M., Clift, J. B., Herrmann, E., & Tomasello, M. (2017). Social disappointment explains chimpanzees' behaviour in the inequity aversion task. *Proceedings of the Royal Society B*, *284*(1861), 20171502. https://doi.org/10.1098/rspb.2017.1502

Erev, I., Shimonovich, D., Schurr, A., & Hertwig, R. (2008). Base rates: How to make the intuitive mind appreciate or neglect them. In *Intuition in judgment and decision making* (pp. 135-148). Erlbaum.

Ermer, E., Cosmides, L., & Tooby, J. (2008). Relative status regulates risky decision making about resources in men: evidence for the co-evolution of motivation and cognition. *Evolution and Human Behavior*, *29*(2), 106–118. https://doi.org/10.1016/j.evolhumbehav.2007.11.002

Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Evans, J. S. B. T. (2008). *Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition* (SSRN Scholarly Paper No. ID 1082085). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=1082085

Evans, T. A., Beran, M. J., Harris, E. H., & Rice, D. F. (2009). Quantity judgments of sequentially presented food items by capuchin monkeys (*Cebus apella*). *Animal Cognition*, *12*(1), 97–105. https://doi.org/10.1007/s10071-008-0174-z

Fabre, J.-M., Caverni, J. P., & Jungermann, H. (1995). Causality Does Influence Conjunctive Probability Judgments If Context and Design Allow for It. *Organizational Behavior and Human Decision Processes*, *63*(1), 1–5. https://doi.org/10.1006/obhd.1995.1056

Fantino, E., Kulik, J., Stolarz-fantino, S., & Wright, W. (1997). The conjunction fallacy: A test of averaging hypotheses. *Psychonomic Bulletin & Review*, *4*(1), 96–101. https://doi.org/10.3758/BF03210779

Feigenson, L. & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, *97*(3), 295–313. https://doi.org/10.1016/j.cognition.2004.09.010

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: object files versus analog magnitudes. *Psychological Science*, *13*(2), 150–156.

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links Between the Intuitive Sense of Number and Formal Mathematics Ability. *Child Development Perspectives*, *7*(2), 74–79. https://doi.org/10.1111/cdep.12019

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. *Journal of Cognitive Neuroscience*, *15*(1), 47–56. https://doi.org/10.1162/089892903321107819

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

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1–17. https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

Fisher, R. A. (1951). The design of experiments 6th ed. *Olivier and Boyd, London, UK*.

Fisk, J. E. (2005). Age and probabilistic reasoning: Biases in conjunctive, disjunctive and Bayesian judgements in early and late adulthood. *Journal of Behavioral Decision Making*, *18*(1), 55–82. https://doi.org/10.1002/bdm.488

Flombaum, J. I., Junge, J. A., & Hauser, M. D. (2005). Rhesus monkeys (*Macaca mulatta*) spontaneously compute addition operations over large numbers. *Cognition*, *97*(3), 315–325. https://doi.org/10.1016/j.cognition.2004.09.004

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

Galbraith, R. C., & Underwood, B. J. (1973). Perceived frequency of concrete and abstract words. *Memory & Cognition*, *1*(1), 56–60. https://doi.org/10.3758/BF03198068

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1–2), 43–74.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65. https://doi.org/10.1016/S1364-6613(99)01424-2

Garland, A., Low, J., & Burns, K. C. (2012). Large quantity discrimination by North Island robins (*Petroica longipes*). *Animal Cognition*, *15*(6), 1129–1140. https://doi.org/10.1007/s10071-012-0537-3

Gibbon, J., & Church, R. M. (1981). Time left: linear versus logarithmic subjective time. *Journal of Experimental Psychology. Animal Behavior Processes*, *7*(2), 87–107.

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. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In *Subjective probability* (pp. 129-161). Wiley.

Gigerenzer, G. (1996). On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky. *Psychological Review*, *103*(3), 592–596.

Gigerenzer, G. (2008). Why Heuristics Work. *Perspectives on Psychological Science*, *3*(1), 20–29. https://doi.org/10.1111/j.1745-6916.2008.00058.x

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Hoffrage, U. (1995). How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review*, *102*(4), 684–704.

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

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. *Unintended thought*, *26*, 481.

Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *The Quarterly Journal of Experimental Psychology*, *64*(11), 2099–2109. https://doi.org/10.1080/17470218.2011.574710

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

Ginosar, Z., & Trope, Y. (1980). The effects of base rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology*, *16*(3), 228–242. https://doi.org/10.1016/0022-1031(80)90066-9

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

Glazko, G. V., & Nei, M. (2003). Estimation of Divergence Times for Major Lineages of Primate Species. *Molecular Biology and Evolution*, *20*(3), 424–434. https://doi.org/10.1093/molbev/msg050

Goldberg, S. (1966). Probability Judgments by Preschool Children: Task Conditions and Performance. *Child Development*, *37*(1), 157–167. https://doi.org/10.2307/1126436

Gómez-Laplaza, L. M., & Gerlai, R. (2011a). Can angelfish (*Pterophyllum scalare*) count? Discrimination between different shoal sizes follows Weber's law. *Animal Cognition*, *14*(1), 1–9. https://doi.org/10.1007/s10071-010-0337-6

Gómez-Laplaza, L. M., & Gerlai, R. (2011b). Spontaneous discrimination of small quantities: shoaling preferences in angelfish (*Pterophyllum scalare*). *Animal Cognition*, *14*(4), 565–574. https://doi.org/10.1007/s10071-011-0392-7

Gómez-Laplaza, L. M., & Gerlai, R. (2013). Quantification abilities in angelfish (*Pterophyllum scalare*): the influence of continuous variables. *Animal Cognition*, *16*(3), 373–383. https://doi.org/10.1007/s10071-012-0578-7

Goodman, N. (1955) Fact, fiction, and forecast. *Harvard University Press*, Cambridge, MA

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A Rational Analysis of Rule-Based Concept Learning. *Cognitive Science*, *32*(1), 108–154. https://doi.org/10.1080/03640210701802071

Gopnik, A. (2014) The surprising probability gurus wearing diapers, The Wall Street 377 Journal, https://www.wsj.com/articles/the-surprising-probability-gurus-wearing-diapers-1389401010

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620–629.

Graf, V., Bullock, D. H., & Bitterman, M. E. (1964). Further Experiments on Probability-Matching in the Pigeon1. *Journal of the Experimental Analysis of Behavior*, *7*(2), 151–157. https://doi.org/10.1901/jeab.1964.7-151

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004

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

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-Based Causal Induction. *Psychological Review*, *116*(4), 661–716. https://doi.org/10.1037/a0017201

Griffiths, T., Sanborn, A., Canini, K., & Navarro, D. (2008). Categorization as nonparametric Bayesian. *The probabilistic mind: Prospects for Bayesian cognitive science*, 303-328.

Groves, C. (2005). Order Primates. In: Wilson, D. E., & Reeder, D. M. (Eds.). *Mammal species of the world: a taxonomic and geographic reference* (Vol. 1). JHU Press.

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

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*, 201003095. https://doi.org/10.1073/pnas.1003095107

Halberda, J., & Feigenson, L. (2008). Developmental Change in the Acuity of the "Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-Year-Olds and Adults. *Developmental Psychology*, *44*(5), 1457–1465. https://doi.org/10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences*, *109*(28), 11116–11120. https://doi.org/10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668. https://doi.org/10.1038/nature07246

Hammerstein, P., & Stevens, J. R. (Eds.). (2012). *Evolution and the mechanisms of decision making* (Vol. 11). Cambridge, MA: MIT Press.

Hanus, D., & Call, J. (2007). Discrete quantity judgments in the great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus*): the effect of presenting whole sets versus item-by-item. *Journal of Comparative Psychology*, *121*(3), 241–249. https://doi.org/10.1037/0735-7036.121.3.241

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

Hara, Y., Imanishi, T., & Satta, Y. (2012). Reconstructing the Demographic History of the Human Lineage Using Whole-Genome Sequences from Human and Three Great Apes. *Genome Biology and Evolution*, *4*(11), 1133–1145. https://doi.org/10.1093/gbe/evs075

Hare, B. (2001). Can competitive paradigms increase the validity of experiments on primate social cognition? *Animal Cognition*, *4*(3–4), 269–280. https://doi.org/10.1007/s100710100084

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785. https://doi.org/10.1006/anbe.1999.1377

Hartnett, P., & Gelman, R. (1998). Early understandings of numbers: paths or barriers to the construction of new understandings? *Learning and Instruction*, *8*(4), 341–374. https://doi.org/10.1016/S0959-4752(97)00026-1

Haun, D. B. M., Nawroth, C., & Call, J. (2011). Great Apes' Risk-Taking Strategies in a Decision Making Task. *PLOS ONE*, *6*(12), e28801. https://doi.org/10.1371/journal.pone.0028801

Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: Examinations of content and format. *Cognitive Psychology*, *47*(4), 367–401. https://doi.org/10.1016/S0010-0285(03)00050-1

Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi–free–ranging rhesus monkeys. *Proceedings of the Royal Society of London B: Biological Sciences*, *267*(1445), 829–833. https://doi.org/10.1098/rspb.2000.1078

Hauser, M. D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences*, *93*(4), 1514–1517. https://doi.org/10.1073/pnas.93.4.1514

Heilbronner, S. R., Rosati, A. G., Stevens, J. R., Hare, B., & Hauser, M. D. (2008). A fruit in the hand or two in the bush? Divergent risk preferences in chimpanzees and bonobos. *Biology Letters*, *4*(3), 246–249. https://doi.org/10.1098/rsbl.2008.0081

Henmon, V. A. C. (1906). The time of perception as a measure of differences in sensations (No. 8). *Science Press.*

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

Hicks, L. H. (1956). An analysis of number-concept formation in the Rhesus monkey. *Journal of Comparative and Physiological Psychology*, *49*(3), 212–218.

Holloway, I. D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Developmental Science*, *11*(5), 644–649. https://doi.org/10.1111/j.1467-7687.2008.00712.x

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Hunt, S., Low, J., & Burns, K. (2008). Adaptive numerical competency in a food-hoarding songbird. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1649), 2373–2379. https://doi.org/10.1098/rspb.2008.0702

Hyde, D. C., & Spelke, E. S. (2011). Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*(2), 360–371. https://doi.org/10.1111/j.1467-7687.2010.00987.x

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, *106*(25), 10382–10385. https://doi.org/10.1073/pnas.0812142106

Jaakkola, K., Fellner, W., Erb, L., Rodriguez, M., & Guarino, E. (2005). Understanding of the concept of numerically "less" by bottlenose dolphins (*Tursiops truncatus*). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *119*(3), 296–303. https://doi.org/10.1037/0735-7036.119.3.296

Jacob, S. N., & Nieder, A. (2009). Tuning to non-symbolic proportions in the human frontoparietal cortex. *European Journal of Neuroscience*, *30*(7), 1432–1442. https://doi.org/10.1111/j.1460-9568.2009.06932.x

Jacob, S. N., Vallentin, D., & Nieder, A. (2012). Relating magnitudes: the brain's code for proportions. *Trends in Cognitive Sciences*, *16*(3), 157–166. https://doi.org/10.1016/j.tics.2012.02.002

Jacobs, J. E., & Potenza, M. (1991). The Use of Judgement Heuristics to Make Social and Object Decisions: A Developmental Perspective. *Child Development*, *62*(1), 166–178. https://doi.org/10.1111/j.1467-8624.1991.tb01522.x

Jeong, Y., Levine, S. C., & Huttenlocher, J. (2007). The Development of Proportional Reasoning: Effect of Continuous Versus Discrete Quantities. *Journal of Cognition and Development*, *8*(2), 237–256. https://doi.org/10.1080/15248370701202471

Jordan, K. E., & Brannon, E. M. (2006). Weber's Law influences numerical representations in rhesus macaques (*Macaca mulatta*). *Animal Cognition*, *9*(3), 159–172. https://doi.org/10.1007/s10071-006-0017-8

Jordan, K. E., Brannon, E. M., Logothetis, N. K., & Ghazanfar, A. A. (2005). Monkeys Match the Number of Voices They Hear to the Number of Faces They See. *Current Biology*, *15*(11), 1034–1038. https://doi.org/10.1016/j.cub.2005.04.056

Jordan, K. E., MacLean, E. L., & Brannon, E. M. (2008). Monkeys match and tally quantities across senses. *Cognition*, *108*(3), 617–625. https://doi.org/10.1016/j.cognition.2008.05.006

Kadosh, R. C., & Dowker, A. (2015). *The Oxford Handbook of Numerical Cognition*. Oxford University Press UK.

Kadosh, R. C., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*(2), 132–147. https://doi.org/10.1016/j.pneurobio.2007.11.001

Kahneman, D., & Frederick, S. (2002). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases* (1st ed., pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kahneman, D., & Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness. In *The Concept of Probability in Psychological Experiments* (pp. 25–48). Springer, Dordrecht. https://doi.org/10.1007/978-94-010-2288-0_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. (1996). On the Reality of Cognitive Illusions. *Psychological Review*, *103*(3), 582–591.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–234. https://doi.org/10.1016/j.cognition.2008.08.010

Kant, I. (1781). Kritik der reinen Vernunft [krv]. *Hamburg: Meiner*.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The Discrimination of Visual Number. *The American Journal of Psychology*, *62*(4), 498–525. https://doi.org/10.2307/1418556

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. https://doi.org/10.1111/j.1467-7687.2007.00585.x

Kilian, A., Yaman, S., von Fersen, L., & Güntürkün, O. (2003). A bottlenose dolphin discriminates visual stimuli differing in numerosity. Animal Learning & Behavior, 31(2), 133-142.

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

Kitchen, D. M. (2004). Alpha male black howler monkey responses to loud calls: effect of numeric odds, male companion behaviour and reproductive investment. *Animal Behaviour*, *67*(1), 125–139. https://doi.org/10.1016/j.anbehav.2003.03.007

Koehler, O. (n.d.). „Zähl"-Versuche an einem Kolkraben und Vergleichsversuche an Menschen. *Zeitschrift Für Tierpsychologie*, *5*(3), 575–712. https://doi.org/10.1111/j.1439-0310.1943.tb00665.x

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

Krupenye, C., Rosati, A. G., & Hare, B. (2015). Bonobos and chimpanzees exhibit human-like framing effects. *Biology Letters*, *11*(2), 20140527. https://doi.org/10.1098/rsbl.2014.0527

Krusche, P., Uller, C., & Dicke, U. (2010). Quantity discrimination in salamanders. *Journal of Experimental Biology*, *213*(11), 1822–1828. https://doi.org/10.1242/jeb.039297

Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General*, *136*(3), 430–450. https://doi.org/10.1037/0096-3445.136.3.430

Kushnir, T., & Gopnik, A. (2005). Young Children Infer Causal Strength From Probabilities and Interventions. *Psychological Science*, *16*(9), 678–683. https://doi.org/10.1111/j.1467-9280.2005.01595.x

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *43*(1), 186–196. https://doi.org/10.1037/0012-1649.43.1.186

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

Lakshminarayanan, V. R., Chen, M. K., & Santos, L. R. (2011). The evolution of decision-making under risk: Framing effects in monkey risk preferences. *Journal of Experimental Social Psychology*, *47*(3), 689–693. https://doi.org/10.1016/j.jesp.2010.12.011

Lambrechts, A., Walsh, V., & Wassenhove, V. van. (2013). Evidence Accumulation in the Magnitude System. *PLOS ONE*, *8*(12), e82122. https://doi.org/10.1371/journal.pone.0082122

Lany, J., & Saffran, J. R. (2010). From Statistics to Meaning: Infants' Acquisition of Lexical Categories. *Psychological Science*, *21*(2), 284–291. https://doi.org/10.1177/0956797609358570

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

Lemaître, J. F., Ramm, S. A., Hurst, J. L., & Stockley, P. (2011). Social cues of sperm competition influence accessory reproductive gland size in a promiscuous mammal. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1709), 1171-1176.

Lewis, K. P., Jaffe, S., & Brannon, E. M. (2005). Analog number representations in mongoose lemurs (*Eulemur mongoz*): evidence from a search task. *Animal Cognition*, *8*(4), 247–252. https://doi.org/10.1007/s10071-004-0251-x

Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science*, *13*(6), 900–906. https://doi.org/10.1111/j.1467-7687.2009.00948.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013a). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology*, *116*(4), 829–838. https://doi.org/10.1016/j.jecp.2013.08.003

Libertus, M. E., Feigenson, L., & Halberda, J. (2013b). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences*, *25*, 126–133. https://doi.org/10.1016/j.lindif.2013.02.001

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, *141*(3), 373–379. https://doi.org/10.1016/j.actpsy.2012.09.009

Libertus, M. E., Pruitt, L. B., Woldorff, M. G., & Brannon, E. M. (2009). Induced alpha-band oscillations reflect ratio-dependent number discrimination in the infant brain. *Journal of Cognitive Neuroscience*, *21*(12), 2398–2406. https://doi.org/10.1162/jocn.2008.21162

Lima, S. L. (1984). Downy Woodpecker Foraging Behavior: Efficient Sampling in Simple Stochastic Environments. *Ecology*, *65*(1), 166–174. https://doi.org/10.2307/1939468

Lima, S. L. (1985). Sampling behavior of starlings foraging in simple patchy environments. *Behavioral Ecology and Sociobiology*, *16*(2), 135–142. https://doi.org/10.1007/BF00295147

Lippmann, W. 1922. Public opinion. New York: Macmillan.

Lipton, J. S., & Spelke, E. S. (2003). Origins of Number Sense: Large-Number Discrimination in Human Infants. *Psychological Science*, *14*(5), 396–401. https://doi.org/10.1111/1467-9280.01453

Longo, N. (1964). Probability-Learning and Habit-Reversal in the Cockroach. *The American Journal of Psychology*, *77*(1), 29–41. https://doi.org/10.2307/1419269

Lourenco, S. F. (2015). Chapter 6 - On the Relation between Numerical and Non-Numerical Magnitudes: Evidence for a General Magnitude System. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Mathematical Cognition and Learning* (Vol. 1, pp. 145–174). Elsevier. https://doi.org/10.1016/B978-0-12-420133-0.00006-5

Lourenco, S. F., & Bonny, J. W. (2014). Representations of numerical and non-numerical magnitude both contribute to mathematical competence in children. *Developmental Science*, *20*(4), e12418. https://doi.org/10.1111/desc.12418

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences*, *109*(46), 18737–18742. https://doi.org/10.1073/pnas.1207212109

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

Luttbeg, B., & Warner, R. R. (1999). Reproductive decision-making by female peacock wrasses: flexible versus fixed behavioral rules in variable environments. *Behavioral Ecology*, *10*(6), 666–674. https://doi.org/10.1093/beheco/10.6.666

Ma, L. 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

Macdonald, R. R., & Gilhooly, K. J. (1990). More about linda or conjunctions in context. *European Journal of Cognitive Psychology*, *2*(1), 57–70. https://doi.org/10.1080/09541449008406197

Mandler, G., & Shebo, B. J. (1982). Subitizing: An Analysis of its Component Processes. *Journal of Experimental Psychology: General*, *111*(1), 1–22.

Manis, M., Dovalina, I., Avis, N. E., & Cardoze, S. (1980). Base rates can affect individual predictions. *Journal of Personality and Social Psychology*, *38*(2), 231–248. https://doi.org/10.1037/0022-3514.38.2.231

Markman, E. M. (1990). Constraints Children Place on Word Meanings. *Cognitive Science*, *14*(1), 57–77. https://doi.org/10.1207/s15516709cog1401_4

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, *315*(6014), 57–59. https://doi.org/10.1038/315057a0

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' Precision of the Approximate Number System Predicts Later School Mathematics Performance. *PLOS ONE*, *6*(9), e23749. https://doi.org/10.1371/journal.pone.0023749

McComb, K., Packer, C., & Pusey, A. (1994). Roaring and Numerical Assessment in Contests Between Groups of Female Lions, Panthera leo. *Animal Behaviour*, *47*(2), 379–387. https://doi.org/10.1006/anbe.1994.1052

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science*, *18*(8), 740–745. https://doi.org/10.1111/j.1467-9280.2007.01969.x

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

McNamara, J. M., Green, R. F., & Olsson, O. (2006). Bayes' theorem and its applications in animal behaviour. *Oikos*, *112*(2), 243–251. https://doi.org/10.1111/j.0030-1299.2006.14228.x

Mechner, F. (1958). Probability Relations Within Response Sequences Under Ratio Reinforcement1. *Journal of the Experimental Analysis of Behavior*, *1*(2), 109–121. https://doi.org/10.1901/jeab.1958.1-109

Mechner, F., & Guevrekian, L. (1962). Effects of deprivation upon counting and timing in rats. *Journal of the Experimental Analysis of Behavior*, *5*(4), 463–466. https://doi.org/10.1901/jeab.1962.5-463

Melis, A. P., Hare, B., & Tomasello, M. (2006). Chimpanzees recruit the best collaborators. *Science*, *311*(5765), 1297–1300. https://doi.org/10.1126/science.1123007

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do Frequency Representations Eliminate Conjunction Effects? An Exercise in Adversarial Collaboration. *Psychological Science*, *12*(4), 269–275. https://doi.org/10.1111/1467-9280.00350

Menzel, E. W. (1961). Perception of food size in the chimpanzee. *Journal of Comparative and Physiological Psychology*, *54*(5), 588.

Menzel, E. W., & Davenport, R. K. (1962). The effects of stimulus presentation variable upon chimpanzee's selection of food by size. *Journal of Comparative and Physiological Psychology*, *55*(2), 235-239.

Meyer, D. R. (1960). The effects of differential probabilities of reinforcement on discrimination learning by monkeys. *Journal of Comparative and Physiological Psychology*, *53*(2), 173-175.

Mitani, J. C. (2009). Cooperation and Competition in Chimpanzees: Current Understanding and Future Challenges. *Evolutionary Anthropology*, *18*(5), 215–227. https://doi.org/10.1002/evan.20229

Muller, M. N., & Mitani, J. C. (2005). Conflict and cooperation in wild chimpanzees. *Advances in the Study of Behavior, Vol 35*, *35*, 275–331. https://doi.org/10.1016/S0065-3454(05)35007-8

Muncer, S. J. (1983). "Conservations" with a chimpanzee. *Developmental Psychobiology*, *16*(1), 1–11. https://doi.org/10.1002/dev.420160102

Murofushi, K. (1997). Numerical matching behavior by a chimpanzee (*Pan troglodytes*): Subitizing and analogue magnitude estimation. *Japanese Psychological Research*, *39*, 236–252.

Nater, A., Mattle-Greminger, M. P., Nurcahyo, A., Nowak, M. G., Manuel, M. de, Desai, T., … Krützen, M. (2017). Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species. *Current Biology*, *27*(22), 3487-3498.e10. https://doi.org/10.1016/j.cub.2017.09.047

Navarro, D. J. (2006). From natural kinds to complex categories. *Proceedings of the 28th Annual Conference of the Cognitive Science Society (pp. 621–626). Mahwah*. Erlbaum.

Newcombe, N., Huttenlocher, J., & Learmonth, A. (1999). Infants' coding of location in continuous space. *Infant Behavior and Development*, *22*(4), 483–510. https://doi.org/10.1016/S0163-6383(00)00011-4

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience*, *6*(3), 177–190. https://doi.org/10.1038/nrn1626

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences*, *109*(29), 11860–11865. https://doi.org/10.1073/pnas.1204580109

Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. *Annual Review of Neuroscience*, *32*, 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nieder, A., Diester, I., & Tudusciuc, O. (2006). Temporal and Spatial Enumeration Processes in the Primate Parietal Cortex. *Science*, *313*(5792), 1431–1435. https://doi.org/10.1126/science.1130308

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the Quantity of Visual Items in the Primate Prefrontal Cortex. *Science*, *297*(5587), 1708–1711. https://doi.org/10.1126/science.1072493

Nieder, A., & Merten, K. (2007). A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *The Journal of Neuroscience,* 27, 5986-5993.

Nieder, A., & Miller, E. K. (2003). Coding of Cognitive Magnitude: Compressed Scaling of Numerical Information in the Primate Prefrontal Cortex. *Neuron*, *37*(1), 149–157. https://doi.org/10.1016/S0896-6273(02)01144-3

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7457–7462. https://doi.org/10.1073/pnas.0402239101

Nisbett, R. E., & Borgida, E. (1975). Attribution and the Psychology of Prediction. *Journal of Personality and Social Psychology*, *32*(5), 932.

Nisbett, R. E., & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Prentice-Hall.

Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology*, *49*(6), 1103–1112. https://doi.org/10.1037/a0029472

Pahl, M., Si, A., & Zhang, S. (2013). Numerical Cognition in Bees and Other Insects. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00162

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

Pelé, M., Broihanne, M. H., Thierry, B., Call, J., & Dufour, V. (2014). To bet or not to bet? Decision-making under risk in non-human primates. *Journal of Risk and Uncertainty*, *49*(2), 141–166. https://doi.org/10.1007/s11166-014-9202-3

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*(2), 109–130. https://doi.org/10.1017/S0140525X08003543

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2012). Are we good at detecting conflict during reasoning? *Cognition*, *124*(1), 101–106. https://doi.org/10.1016/j.cognition.2012.04.004

Pepperberg, I. M. (2006). Grey parrot numerical competence: a review. *Animal Cognition*, *9*(4), 377–391. https://doi.org/10.1007/s10071-006-0034-7

Pepperberg, I. M. (1987). Evidence for Conceptual Quantitative Abilities in the African Grey Parrot: Labeling of Cardinal Sets. *Ethology*, *75*(1), 37–61. https://doi.org/10.1111/j.1439-0310.1987.tb00641.x

Perdue, B. M., Talbot, C. F., Stone, A. M., & Beran, M. J. (2012). Putting the elephant back in the herd: elephant relative quantity judgments match those of other species. *Animal Cognition*, *15*(5), 955–961. https://doi.org/10.1007/s10071-012-0521-y

Pérez-Escudero, A., & Polavieja, G. G. de. (2011). Collective Animal Behavior from Bayesian Estimation and Probability Matching. *PLOS Computational Biology*, *7*(11), e1002282. https://doi.org/10.1371/journal.pcbi.1002282

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

Petrillo, F. D., Ventricelli, M., Ponsi, G., & Addessi, E. (2015). Do tufted capuchin monkeys play the odds? Flexible risk preferences in *Sapajus spp*. *Animal Cognition*, *18*(1), 119–130. https://doi.org/10.1007/s10071-014-0783-7

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. New York,: Norton.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus. *Neuron*, *44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Piffer, L., Agrillo, C., & Hyde, D. C. (2012). Small and large number discrimination in guppies. *Animal Cognition*, *15*(2), 215–221. https://doi.org/10.1007/s10071-011-0447-9

Pighin, S., Tentori, K., & Girotto, V. (2017). Another chance for good reasoning. *Psychonomic Bulletin & Review*, *24*(6), 1995–2002. https://doi.org/10.3758/s13423-017-1252-5

Politzer, G., & Noveck, I. A. (1991). Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research*, *20*(2), 83–103. https://doi.org/10.1007/BF01067877

Proctor, D., Williamson, R. A., Latzman, R. D., Waal, F. B. M. de, & Brosnan, S. F. (2014). Gambling primates: reactions to a modified Iowa Gambling Task in humans, chimpanzees and capuchin monkeys. *Animal Cognition*, *17*(4), 983–995. https://doi.org/10.1007/s10071-014-0730-7

Rakoczy, H., Cluver, A., Saucke, L., Stoffregen, N., Grabener, 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

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

Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: a review of data and theories. In G. Wright, & P. Ayton (Eds.), *Subjective probability* (pp. 239–272). New York, New York: Wiley.

Roberts, W. A. (2006). Evidence that pigeons represent both time and number on a logarithmic scale. *Behavioural Processes*, *72*(3), 207–214. https://doi.org/10.1016/j.beproc.2006.03.002

Roberts, W. A., MacDonald, H., & Lo, K. H. (2018). Pigeons play the percentages: computation of probability in a bird. *Animal Cognition*, *21*(4), 575–581. https://doi.org/10.1007/s10071-018-1192-0

Ross, G. R. T. (1908). Aristotle de Anima.

Rosati, A. G., & Hare, B. (2012). Decision making across social contexts: competition increases preferences for risk in chimpanzees and bonobos. *Animal Behaviour*, *84*(4), 869–879. https://doi.org/10.1016/j.anbehav.2012.07.010

Rosati, A. G., & Hare, B. (2013). Chimpanzees and Bonobos Exhibit Emotional Responses to Decision Outcomes. *PLOS ONE*, *8*(5), e63058. https://doi.org/10.1371/journal.pone.0063058

Rugani, R., Cavazzana, A., Vallortigara, G., & Regolin, L. (2013). One, two, three, four, or is there something more? Numerical discrimination in day-old domestic chicks. *Animal Cognition*, *16*(4), 557–564. https://doi.org/10.1007/s10071-012-0593-8

Rugani, R., Regolin, L., & Vallortigara, G. (2008). Discrimination of small numerosities in young chicks. *Journal of Experimental Psychology. Animal Behavior Processes*, *34*(3), 388–399. https://doi.org/10.1037/0097-7403.34.3.388

Rumbaugh, D. M., Savage-Rumbaugh, E. S., & Pate, J. L. (1988). Addendum to "Summation in the chimpanzee (*Pan troglodytes*)." *Journal of Experimental Psychology. Animal Behavior Processes*, *14*(1), 118–120.

Rumbaugh, D. M., Savage-Rumbaugh, S., & Hegel, M. T. (1987). Summation in the chimpanzee (*Pan troglodytes*). *Journal of Experimental Psychology. Animal Behavior Processes*, *13*(2), 107–115.

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

Sanborn, A., Chater, N., & Heller, K. A. (2009). Hierarchical Learning of Dimensional Biases in Human Categorization. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 727–735). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/3849-hierarchical-learning-of-dimensional-biases-in-human-categorization.pdf

Schmelz, M., & Call, J. (2016). The psychology of primate cooperation and competition: a call for realigning research agendas. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150067. https://doi.org/10.1098/rstb.2015.0067

Schmelz, M., Call, J., & Tomasello, M. (2011). Chimpanzees know that others make inferences. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 3077–3079. https://doi.org/10.1073/pnas.1000469108

Schmelz, M., Call, J., & Tomasello, M. (2013). Chimpanzees predict that a competitor's preference will match their own. *Biology Letters*, *9*(1). https://doi.org/10.1098/rsbl.2012.0829

Schmitt, V., & Fischer, J. (2011). Representational format determines numerical competence in monkeys. *Nature Communications*, *2*, 257. https://doi.org/10.1038/ncomms1262

Schmitt, V., Kröger, I., Zinner, D., Call, J., & Fischer, J. (2013). Monkeys perform as well as apes and humans in a size discrimination task. *Animal Cognition*, *16*(5), 829–838. https://doi.org/10.1007/s10071-013-0616-0

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, *43*(5), 1124–1139. https://doi.org/10.1037/0012-1649.43.5.1124

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233–250. https://doi.org/10.1002/bdm.413

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. *Rational models of cognition*, 73-95.

Shinskey, J. L., & Munakata, Y. (2005). Familiarity breeds searching - Infants reverse their novelty preferences when reaching for hidden objects. *Psychological Science*, *16*(8), 596–600. https://doi.org/10.1111/j.1467-9280.2005.01581.x

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: the new frontier for theories of numerical development. *Trends in Cognitive Sciences*, *17*(1), 13–19. https://doi.org/10.1016/j.tics.2012.11.004

Simon, H. A. (1979). Rational Decision Making in Business Organizations. *The American Economic Review*, *69*(4), 493–513.

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, *91*(2), 296–309. https://doi.org/10.1016/S0749-5978(03)00021-9

Smedt, B. D., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

Smith L. B. (2001). How domain-general processes may create domain-specific biases. *Language Acquisition and Conceptual Development*. Retrieved from https://ci.nii.ac.jp/naid/10015080787/

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name Learning Provides On-the-Job Training for Attention. *Psychological Science*, *13*(1), 13–19. https://doi.org/10.1111/1467-9280.00403

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*(3), 303–333. https://doi.org/10.1207/s15516709cog2803_1

Spelke, E. (1994). Initial knowledge: six suggestions. *Cognition*, *50*(1), 431–445. https://doi.org/10.1016/0010-0277(94)90039-6

Spelke, E. S. (2003). What Makes Us Smart? Core Knowledge and Natural Language. In D. Getner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 277–311). MIT Press.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of Knowledge. *Psychological Review*, *99*(4), 605–632.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Spinillo, A. G., & Bryant, P. E. (1999). Proportional Reasoning in Young Children: Part-Part Comparisons about Continuous and Discontinuous Quantity. *Mathematical Cognition*, *5*(2), 181–197. https://doi.org/10.1080/135467999387298

Srinivasan, M., & Carey, S. (2010). The long and the short of it: On the nature and origin of functional overlap between representations of space and time. *Cognition*, *116*(2), 217–241. https://doi.org/10.1016/j.cognition.2010.05.005

Staddon, J. E. R. (1988). Learning as inference. *Evolution and learning*, 59-77.

Stancher, G., Rugani, R., Regolin, L., & Vallortigara, G. (2015). Numerical discrimination by frogs (*Bombina orientalis*). *Animal Cognition*, *18*(1), 219–229. https://doi.org/10.1007/s10071-014-0791-7

Stanovich, K. E., & West, R. F. (1998). Individual Differences in Framing and Conjunction Effects. *Thinking & Reasoning*, *4*(4), 289–317. https://doi.org/10.1080/135467898394094

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18116–18120. https://doi.org/10.1073/pnas.1302751110

Stephens, D. W., & Krebs, J. R. (1986). Foraging theory. *Monographs in Behavior and Ecology (USA)*. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=US8830566

Steyvers, M., Griffiths, T. L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Sciences*, *10*(7), 327–334. https://doi.org/10.1016/j.tics.2006.05.005

Suda, C., & Call, J. (2004). Piagetian Liquid Conservation in the Great Apes (*Pan paniscus, Pan troglodytes, and Pongo pygmaeus*). *Journal of Comparative Psychology*, *118*(3), 265–279. https://doi.org/10.1037/0735-7036.118.3.265

Suda, C., & Call, J. (2005). Piagetian conservation of discrete quantities in bonobos (*Pan paniscus*), chimpanzees (*Pan troglodytes*), and orangutans (*Pongo pygmaeus*). *Animal Cognition*, *8*(4), 220–235. https://doi.org/10.1007/s10071-004-0247-6

Tecwyn, E. C., Denison, S., Messer, E. J., & Buchsbaum, D. (2016). Intuitive probabilistic inference in capuchin monkeys. *Animal Cognition*. https://doi.org/10.1007/s10071-016-1043-9

Teglas, 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

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

Teglas, 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

Teglas, 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

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: a misunderstanding about conjunction? *Cognitive Science*, *28*(3), 467–477. https://doi.org/10.1207/s15516709cog2803_8

Tentori, K., & Crupi, V. (2012). On the conjunction fallacy and the meaning of and, yet again: A reply to Hertwig, Benz, and Krauss (2008). *Cognition*, *122*(2), 123–134. https://doi.org/10.1016/j.cognition.2011.09.002

Thomas, R. K., & Chase, L. (1980). Relative numerousness judgments by squirrel monkeys. *Bulletin of the Psychonomic Society*, *16*(2), 79–82. https://doi.org/10.3758/BF03334444

Tomasello, M., & Call, J. (1997). *Primate cognition*. New York: Oxford University Press.

Tomonaga, M., & Matsuzawa, T. (2002). Enumeration of briefly presented items by the chimpanzee (*Pan troglodytes*) and humans (*Homo sapiens*). *Animal Learning & Behavior*, *30*(2), 143–157. https://doi.org/10.3758/BF03192916

Trick, L., & Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80–102.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological bulletin*, *76*(2), 10

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

Tversky, A., & Kahneman, D. (1993). Probabilistic Reasoning. In A. Goldman (Ed.), *Readings in Philosophy and Cognitive Science* (pp. 43–68). Cambridge: MIT Press.

Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development*, *14*(1), 1–36. https://doi.org/10.1016/S0885-2014(99)80016-1

Uller, C., Jaeger, R., Guidry, G., & Martin, C. (2003). Salamanders (*Plethodon cinereus*) go for more: rudiments of number in an amphibian. *Animal Cognition*, *6*(2), 105–112. https://doi.org/10.1007/s10071-003-0167-x

Uller, C., & Lewis, J. (2009). Horses (*Equus caballus*) select the greater of two quantities in small numerical contrasts. *Animal Cognition*, *12*(5), 733–738. https://doi.org/10.1007/s10071-009-0225-0

Vallentin, D., & Nieder, A. (2008). Behavioral and Prefrontal Representation of Spatial Proportions in the Monkey. *Current Biology*, *18*(18), 1420–1425. https://doi.org/10.1016/j.cub.2008.08.042

Vallentin, D., & Nieder, A. (2010). Representations of visual proportions in the primate posterior parietal and prefrontal cortices. *The European Journal of Neuroscience*, *32*(8), 1380–1387. https://doi.org/10.1111/j.1460-9568.2010.07427.x

Vallortigara, G. (2015). Chapter 2 - Foundations of Number and Space Representations in Non-Human Species. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Mathematical Cognition and Learning* (Vol. 1, pp. 35–66). Elsevier. https://doi.org/10.1016/B978-0-12-420133-0.00002-8

Valone, T. J. (2006). Are animals capable of Bayesian updating? An empirical review. *Oikos*, *112*(2), 252–259. https://doi.org/10.1111/j.0030-1299.2006.13465.x

vanMarle, K., Aw, J., McCrink, K., & Santos, L. R. (2006). How capuchin monkeys (*Cebus apella*) quantify objects and substances. *Journal of Comparative Psychology*, *120*(4), 416–426. https://doi.org/10.1037/0735-7036.120.4.416

vanMarle, K. van, Chu, F. W., Li, Y., & Geary, D. C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Developmental Science*, *17*(4), 492–505. https://doi.org/10.1111/desc.12143

vanMarle, K., & Wynn, K. (2006). Six-month-old infants use analog magnitudes to represent duration. *Developmental Science*, *9*(5), F41–F49. https://doi.org/10.1111/j.1467-7687.2006.00508.x

Vonk, J. (2014). Quantity matching by an orangutan (*Pongo abelii*). *Animal Cognition*, *17*(2), 297–306. https://doi.org/10.1007/s10071-013-0662-7

Vonk, J., & Beran, M. J. (2012). Bears 'count' too: quantity estimation and comparison in black bears, *Ursus americanus*. *Animal Behaviour*, *84*(1), 231–238. https://doi.org/10.1016/j.anbehav.2012.05.001

Von Plato, J. (1994). Creating modern probability: Its mathematics, physics and philosophy in historical perspective. *Cambridge University Press.*

Vulkan, N. (2002). An Economist's Perspective on Probability Matching. *Journal of Economic Surveys*, *14*(1), 101–118. https://doi.org/10.1111/1467-6419.00106

Wedell, D. H., & Moro, R. (2008). Testing boundary conditions for the conjunction fallacy: Effects of response mode, conceptual focus, and problem type. *Cognition*, *107*(1), 105–136. https://doi.org/10.1016/j.cognition.2007.08.003

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

Wells, G. L. (1985). The Conjunction Error and the Representativeness Heuristic. *Social Cognition*, *3*(3), 266–279. https://doi.org/10.1521/soco.1985.3.3.266

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal Counting in Humans: The Psychophysics of Number Representation. *Psychological Science*, *10*(2), 130–137. https://doi.org/10.1111/1467-9280.00120

Wilson, M. L., Britton, N. F., & Franks, N. R. (2002). Chimpanzees and the Mathematics of Battle. *Proceedings of the Royal Society of London B Biological Sciences*, *269*(1496), 1107–1112.

Wilson, W. A., Oscar Jr, M., & Bitterman, M. E. (1964). Probability-learning in the monkey. *Quarterly Journal of Experimental Psychology*, *16*(2), 163–165. https://doi.org/10.1080/17470216408416361

Wodinsky, J., Bitterman, M. E., & Candland, D. K. (1958). Some Comparative Psychology. *American Journal of Psychology; Urbana, Etc.*, *71*(1), 94–110.

Woodruff, G., & Premack, D. (1981). Primative mathematical concepts in the chimpanzee: proportionality and numerosity. *Nature*, *293*(5833), 568–570.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750. https://doi.org/10.1038/358749a0

Wynne, C. D. L. (2004). Animal behaviour: Fair refusal by capuchin monkeys. *Nature*, *428*(6979), 140. https://doi.org/10.1038/428140a

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, *89*(1), B15–B25. https://doi.org/10.1016/S0010-0277(03)00050-7

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 of the United States of America*, *105*(13), 5012–5015. https://doi.org/10.1073/pnas.0704450105

Xu, F, & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. https://doi.org/10.1016/S0010-0277(99)00066-9

Xu, F., & Tenenbaum, J. B. (2007b). Word Learning as Bayesian Inference. *Psychological Review*, *114*(2), 245–272.

Xu, F., & Tenenbaum, J. B. (2007b). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, *10*(3), 288–297. https://doi.org/10.1111/j.1467-7687.2007.00590.x

Yamamoto, S., Humle, T., & Tanaka, M. (2012). Chimpanzees' flexible targeted helping based on an understanding of conspecifics' goals. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3588–3592. https://doi.org/10.1073/pnas.1108517109

Yates, J. F., & Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgment procedures, including "signed summation." *Organizational Behavior and Human Decision Processes*, *37*(2), 230–253. https://doi.org/10.1016/0749-5978(86)90053-1

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

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308. https://doi.org/10.1016/j.tics.2006.05.002

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

Appendix A: Eckert, Rakoczy, & Call (2017)

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)*.* doi:10.1002/ajp.22693

WILEY **PRIMATOLOGY**

Are great apes able to reason from multi-item samples to populations of food items?

Johanna Eckert[1,2] | Hannes Rakoczy[2] | Josep Call[1,3]

[1] Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[2] Department of Developmental Psychology, University of Goettingen, Goettingen, Germany

[3] School of Psychology and Neuroscience, University of St Andrews, St Andrews, United Kingdom

Correspondence
Johanna Eckert, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig 04103, Germany.
Email: johanna_eckert@eva.mpg.de

Funding information
Deutsche Forschungsgemeinschaft, Grant number: RA 2155/3-1

Inductive learning from limited observations is a cognitive capacity of fundamental importance. In humans, it is underwritten by our intuitive statistics, the ability to draw systematic inferences from populations to randomly drawn samples and vice versa. According to recent research in cognitive development, human intuitive statistics develops early in infancy. Recent work in comparative psychology has produced first evidence for analogous cognitive capacities in great apes who flexibly drew inferences from populations to samples. In the present study, we investigated whether great apes (*Pongo abelii*, *Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*) also draw inductive inferences in the opposite direction, from samples to populations. In two experiments, apes saw an experimenter randomly drawing one multi-item sample from each of two populations of food items. The populations differed in their proportion of preferred to neutral items (24:6 vs. 6:24) but apes saw only the distribution of food items in the samples that reflected the distribution of the respective populations (e.g., 4:1 vs. 1:4). Based on this observation they were then allowed to choose between the two populations. Results show that apes seemed to make inferences from samples to populations and thus chose the population from which the more favorable (4:1) sample was drawn in Experiment 1. In this experiment, the more attractive sample not only contained proportionally but also absolutely more preferred food items than the less attractive sample. Experiment 2, however, revealed that when absolute and relative frequencies were disentangled, apes performed at chance level. Whether these limitations in apes' performance reflect true limits of cognitive competence or merely performance limitations due to accessory task demands is still an open question.

KEYWORDS
comparative cognition, intuitive statistics, non-human primates, numerical cognition, probabilistic reasoning

1 | INTRODUCTION

Making general inferences from limited data is one of the key components of human inductive learning (see e.g., Denison & Xu, 2012; Holland, 1986; Tenenbaum, Griffiths, & Kemp, 2006; Skyrms, 1975). Traditionally, statistical reasoning was deemed to be difficult and error-prone (e.g., Cosmides & Tooby, 1996; Tversky & Kahneman, 1974, 1981) and dependent on language and formal education

(e.g., Piaget & Inhelder, 1975). However, recent studies suggest that even very young human infants have an astonishingly broad understanding of statistical relations: They are able to generalize from small samples to larger populations (Denison, Reed, & Xu, 2013; Xu & Garcia, 2008), make predictions about single event probabilities (e.g., Téglás, Girotto, Gonzalez, & Bonatti, 2007) and use these predictions to guide their search for desired objects (Denison & Xu, 2010b, 2014; Feigenson, Carey, & Hauser, 2002). In one remarkable

study, for example, infants were confronted with two jars containing mixtures of preferred and non-preferred types of candy (Denison & Xu, 2010b). After they had watched the experimenter randomly sampling one piece of candy from each jar and placing it in an occluded cup, most infants searched in the cup that contained a sample from the jar with a higher proportion of their preferred candy (Denison & Xu, 2010b). Hence, infants seem to have used the proportional information provided by the populations to reason about the samples. Moreover, infants can integrate probabilistic information with information from other domains such as intuitive physics or intuitive psychology (Denison, Trikutam, & Xu, 2014; Téglás, Vul, & Girotto, 2011; Denison & Xu 2010a; Xu & Denison 2009). For example, infants understand that a preference of the experimenter for a certain type of object can turn a sampling process into a non-random event. If the same experimenter, however, is blindfolded, infants expect the sampled objects to reflect the proportions within populations (Xu & Denison 2009). These findings imply that at least at the age of 6 months, humans already flexibly use intuitive statistics to predict the outcome of events. Being apparently independent of language or formal education, this raises the question whether these kinds of probabilistic reasoning represent an evolutionary ancient trait that is shared with other species.

Many species are capable of numerical cognition: For example, great apes (e.g., Beran, McIntyre, Garland, & Evans, 2013; Boysen & Berntson, 1989; Call, 2000; Hanus & Call, 2007), old- and new-world monkeys (e.g., Barnard, Hughes, & Gerhardt, 2013; Beran, Evans, Leighty, Harris, & Rice, 2008; Beran & Parrish, 2016), elephants (Perdue, Talbot, Stone, & Beran, 2012), bears (Vonk & Beran, 2012), raccoons (Davis, 1984), dogs (Ward & Smuts, 2007), cats (Pisa & Agrillo, 2009), birds [e.g., Rugani, Cavazzana, Vallortigara, & Regolin, 2013], fish [e.g., Potrich, Sovrano, Stancher, & Vallortigara, 2015], and even insects [bees: Dacke & Srinivasan, 2008; ants: Reznikova & Ryabko, 2011] are able to compare quantities, suggesting that representing numerosity is an evolutionary ancient trait. The practical advantages of such a capacity are obvious: in the context of foraging, for example, comparing quantities is a highly useful tool to identify the most profitable feeding location [see e.g., Farnsworth & Smolinski, 2006 and Hunt, Low, & Burns, 2008 for field experiments on quantity discrimination in a foraging context]. In the context of competition comparing one's own group size with that of a rival group can help to estimate the chances of winning a potential fight [e.g., Benson-Amram, Heinen, Dryer, & Holekamp, 2011; McComb, Packer, & Pusey, 1994; Wilson, Britton, & Franks, 2002]. Chimpanzees, for example, have been found to attack an opponent group only if their own group outnumbers those of their conspecifics by at least 1.5 (Wilson et al., 2002).

Relatedly, one can imagine that in some situations it would make sense for an animal to be able to make probability judgments instead of straightforward quantity comparisons. Efficient foraging, for instance, requires an individual to search for food in locations that most likely provide the best payoff in relation to foraging time (Geary, Berch, & Mann Koepke, 2015; for a review about optimal foraging theory see example, Hamilton, 2010). One possibility to identify the best payoff

per time unit is to use the relative frequency of past successes in a feeding location. Imagine a group of chimpanzees that has to decide in the morning in which direction to go: Either toward feeding ground A or toward feeding ground B. The apes might want to compare the proportion of times they visited each feeding ground and obtained a sufficient amount of food instead of simply comparing the absolute number of times they were successful in each location. Hence, non-human animals could clearly benefit from an ability exceeding a mere estimation of absolute or relative numerosity, namely a sense for probabilistic relations, that is, intuitive statistics. Future research will need to investigate both when and due to which selection pressures intuitive statistics evolved.

A recent comparative study investigated intuitive statistical abilities in non-human great apes with the same kinds of methods used in infancy research (Rakoczy, Clüver, & Saucke, 2014). Individuals of four great ape species were presented with two populations of food items. Both populations consisted of the same two types of food (one type clearly preferred over the other) but with different relative frequency distributions. The experimenter drew a one-object-sample from each population and gave the subject a choice between the two hidden samples. Hence, subjects had to infer which population was more likely to yield a preferred food item as a sample. Interestingly, individuals of all tested great ape species were able to form correct expectations about the probability of the sampling events, even when absolute and relative frequencies within the populations were disentangled. Apes' inferences were, therefore, not only based on information about absolute frequency, but instead they were truly based on probabilistic information. Most recently, another representative of the primate order was tested in the same paradigm: Capuchin monkeys Sapajus sp (Tecwyn, Denison, Messer, & Buchsbaum, 2016). In a series of four experiments, the monkeys were allowed to choose between the randomly drawn samples of two populations of food items with different proportions of preferred and non-preferred food. Results revealed that a few individuals might have drawn probabilistic inferences based on proportional information (control conditions excluded the usage of simpler choice heuristics). However, monkey's performance in a baseline control condition was unexpectedly low, questioning whether they truly fully understood the procedure. It remains, therefore, an open question whether primates other than great apes are capable of intuitive statistics.

Although the findings of Rakoczy et al. (2014) raised the possibility that apes and human infants may operate with the same cognitive capacities for intuitive statistics, it leaves many open questions for future research. One fundamentally important question is whether apes' intuitive statistics reveal the same kinds of flexibility and generality as those found in human infants. In particular, does their ability to form expectations about samples randomly drawn from populations (inference population > sample; Rakoczy et al., 2014) extend to the ability to reason from a given sample to the corresponding population (inference sample > population)?

In human infants, this question was addressed using the violation of expectation (VOE) looking-time paradigm (Denison et al., 2013; Xu & Garcia, 2008). In one study (Xu & Garcia 2008) 8-month-old infants

were presented with boxes containing populations of red and white Ping-Pong balls. The distribution of red to white balls was either 9:1 or 1:9. During test-trials, the box containing one of the two populations of Ping-Pong balls was covered and the infants watched the experimenter drawing (apparently randomly) a sample of either four red and one white Ping-Pong balls or one red and four white Ping-Pong balls. Subsequently, the experimenter removed the cover of the box she had drawn from and revealed the population. Infants looked longer at the "mostly red"-sample when it was drawn from the "mostly white" population (unexpected) than when it was drawn from the "mostly red" population (expected). The analogue was true for the "mostly white"-sample. In a control condition it could be ruled out that infants simply reacted to the perceptual mismatch between sample and population: Instead of drawing the balls as samples from the box, the experimenter pulled them out of her pocked and placed them next to the box, resulting in equal looking-times at both matched and mismatched outcomes. This implies that, confronted with a sample, infants were able to make inferences about the associated population. Applying the same paradigm, a second study (Denison et al., 2013) showed that even 6-month-old infants had intuitions about relationships between samples and populations, suggesting that the ability to make inferences based on samples develops very early in human ontogeny. The results of these two infant studies (Denison et al., 2013; Xu & Garcia, 2008) together with the findings of the first study on intuitive statistics in great apes (Rakoczy et al., 2014) may indicate that the capacity of making inductive inferences is shared with other non-human species.

In the current study we tested great apes' ability to reason from samples to populations. Combining the methodology of Rakoczy et al. (2014) and Xu & Garcia (2008), we confronted apes with two covered containers holding populations of food items that differed in their proportion of preferred to neutral food (24:6 vs. 6:24). In two experiments, the apes witnessed the experimenter drawing one multi-item sample from each population. The distribution within the multi-item samples reflected the distribution of the respective populations (e.g., 4:1 vs. 1:4). Based on the observation of these representative samples, the apes were allowed to choose between the two covered populations. Hence, to receive the more favorable population, they were required to use proportional information provided by the samples. In Experiment 1 we tested whether apes were able to reason from samples to populations. In the crucial 4:1 versus 1:4 Test condition, apes were confronted with two populations (24:6 vs. 6:24) and watched the samples 4:1 versus 1:4 being drawn from them. Two control conditions were designed to rule out alternative explanations. In Control condition 1 apes did not see the available populations beforehand. This manipulation tested whether the information provided by the samples alone was sufficient for the apes to infer about the distribution within the populations. In Control condition 2, the samples were not re-inserted into the populations after the drawing process. This manipulation tested whether apes' success in previous conditions might have reflected a tendency to choose the population where the more favorable sample was inserted, without necessarily having to reason about the drawing process. In Experiment

2, apes were tested in two further Test conditions in which absolute and relative frequencies of preferred food items within samples were disentangled: In the 2:1 versus 4:8 Test condition, the absolute number of preferred food items was lower in the sample drawn from the more favorable population and therefore, misleading. In the 4:1 versus 4:8 Test condition, the absolute number of preferred food items was the same in both samples and therefore inconclusive. Hence, to receive the more favorable population in Experiment 2, apes had to take into account proportions, rather than absolute numbers.

## 2 | EXPERIMENT 1: CAN APES REASON FROM SAMPLES TO POPULATIONS?

In this experiment we sought to investigate whether apes were able to reason from multi-item samples to populations. In the Test condition, the experimenter presented the apes with two covered containers holding populations of food items (24:6 vs. 6:24). After watching representative samples being drawn from those populations (4:1 vs. 1:4), subjects were allowed to choose between the two containers. Two control conditions (1) tested whether apes inferred from the samples alone which distribution the populations had and (2) ruled out that subjects used the simple heuristic of choosing the container where the more attractive sample was inserted after the sampling process (see Figure 1 for an illustration of the different Test conditions). Based on the results of Rakoczy et al. (2014) we expected no inter-specific differences.

2.1 | Methods

2.1.1 | Subjects

Twenty-six individuals (female $N = 20$) of four great ape species participated: Gorillas (*Gorilla gorilla*, $N = 4$), Bonobos (*Pan paniscus*, $N = 6$), Chimpanzees (*Pan troglodytes*, $N = 10$), and Orangutans (*Pongo abelii*, $N = 6$). One further chimpanzee was tested but excluded from data analysis since he did not complete all sessions due to lack of motivation. Subjects were housed at the Wolfgang Koehler Primate Research Center (WKPRC) in the Leipzig Zoo and were tested between November 2014 and September 2015. Their age ranged between six and 48 years (Mean = 18 years) and about 25% were hand-reared. The remaining 75% were mother-reared (see Table 1 for more subject information). All apes were already experienced in participating in cognitive tasks with food-rewards as reinforcement. To control for potential order effects, 15 of the subjects underwent Experiment 1 first and then proceeded to Experiment 2, the remaining 11 subjects experienced Experiment 2 first and were tested in Experiment 1 afterwards (see Figure 2).

The study was ethically approved by an internal committee at the Max Planck Institute for Evolutionary Anthropology. Research and animal husbandry comply with the "EAZA Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria," the "EEP Bonobo Husbandry Manual," the "WAZA Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums" and the "Guidelines for the Treatment of Animals in Behavioral Research and
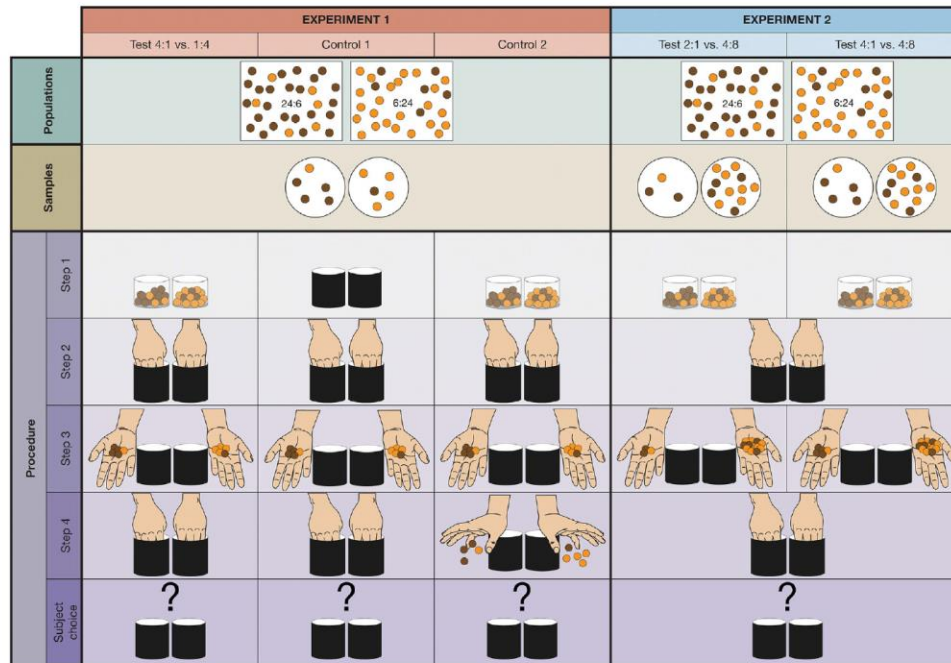
FIGURE 1 Schematic illustration of the different test conditions. Experiment 1 comprised one test and two control conditions. In Control 1, the populations were not shown prior to the sampling process. In Control 2, the samples were not replaced after sampling. The populations (24:6 vs. 6:24) as well as the samples (4:1 vs. 1:4) remained the same in all conditions of Experiment 1. Experiment 2 comprised two test conditions with the samples 2:1 versus 4:8 and 4:1 versus 4:8, respectively. The procedure and the populations remained the same as in the test condition of Experiment 1

Teaching" of the Association for the Study of Animal Behavior (ASAB). This research adhered to the American Society of Primatologists principles for the ethical treatment of primates.

2.1.2 | Materials

Subjects were tested individually in their sleeping cages or in special test cages. A Plexiglas panel fitted on the cage mesh separated ape and experimenter. The panel had two small holes (ø 2 cm; distance between holes 59 cm) through which subjects could insert a finger to indicate a choice. Perpendicular to the Plexiglas panel, a sliding table (45 × 79 cm) was mounted on the cage and could be moved both toward the subject and the experimenter. Underneath the table there were two small concealed compartments in which food items could be hidden prior to each test trial without the subject noticing it. To prevent subjects from watching, for example, preparation of a trial, a screen (27 × 79 cm) could be fixed via metal brackets at the end of the table closer to the ape's side. During test trials, apes were presented with two transparent Plexiglas containers (ø 8 cm), each containing a population of food items, namely pieces of fruit pellets and pieces of carrots of roughly equal shape and size. The containers could be covered with opaque occluders of the same diameter that prevented subjects from seeing the content of the containers (see Figure 3 for an illustration of the setup).

2.1.3 | Design and procedure

Before the actual test started, subjects underwent a familiarization session. Subsequently, we carried out one test and two control conditions to investigate whether apes were able to reason from multi-item samples to populations. All conditions consisted of 12 test trials, divided into three sessions. Each session started with two preference trials with single pellet and carrot pieces (see Preference trials section). Thus, each session consisted of two preference trials and four test trials.

Familiarization

Each subject that had not experienced Experiment 2 before received one session with six trials of familiarization. In this session, the ape was confronted with one transparent container holding a population of carrot and pellet pieces (distribution 12:12). The experimenter presented the container to the ape, shook it several times to give a good overview of the population and subsequently placed it in the center of the sliding table.

TABLE 1 Detailed description of animals participating in this study

Subject	Species	Sex	Age	Rearing history	Start with experiment	Order of conditions	Drop out, reason
Fimi	Bonobo	F	7	Mother	2	Second	Yes, "covered population" follow-up
Gemena	Bonobo	F	9	Mother	2	Second	No
Kuno	Bonobo	M	18	Nursery	2	Second	No
Lexi	Bonobo	F	15	Nursery	1	First	No
Luiza	Bonobo	F	10	Mother	1	First	No
Yasa	Bonobo	F	18	Mother	1	First	No
Alexandra	Chimpanzee	F	15	Nursery	2	Second	No
Bangolo	Chimpanzee	M	6	Mother	1	First	Yes, lack of motivation
Daza	Chimpanzee	F	28	Unknown	2	Second	Yes, "open population" follow-up
Frodo	Chimpanzee	M	21	Mother	1	First	No
Jahaga	Chimpanzee	F	22	Mother	2	Second	No
Jeudi	Chimpanzee	F	48	Unknown	2	Second	Yes, "covered population" follow-up
Kara	Chimpanzee	F	9	Mother	2	Second	No
Lobo	Chimpanzee	M	10	Mother	1	First	No
Riet	Chimpanzee	F	37	Nursery	1	First	No
Robert	Chimpanzee	M	38	Nursery	2	Second	Yes, lack of motivation
Sandra	Chimpanzee	F	21	Mother	1	First	No
Tai	Chimpanzee	F	12	Mother	1	First	No
Abeeku	Gorilla	M	15	Mother	1	First	Yes, "covered population" follow-up
Kibara	Gorilla	F	10	Mother	1	First	No
Kumili	Gorilla	F	10	Mother	1	First	Yes, "covered population" follow-up
Viringika	Gorilla	F	19	Mother	1	First	No
Bimbo	Orangutan	M	34	Nursery	2	Second	No
Dokana	Orangutan	F	26	Mother	1	First	No
Padana	Orangutan	F	17	Mother	1	First	No
Pini	Orangutan	F	27	Mother	2	Second	No
Raja	Orangutan	F	11	Mother	2	Second	No
Suaq	Orangutan	M	6	Mother	1	First	No

During the first three trials the subject watched the experimenter drawing a random sample (three to five items) out of the population, presenting it on the palm of the hand, and re-inserting it into the container. After that, the experimenter moved the container to the edge of the table and pushed the sliding table forward, so that the ape could point to the container. Subsequently, the subject received the content of the container as reward. During the last three trials of familiarization, the procedure was the same as explained above, but this time the container was placed in an opaque occluder after the ape had seen the population. Thus, the subject did not see the population during the sampling process and when pointing to it. The familiarization should ensure that subjects were familiar with the material and that they understood that "inserting a hand in an occluded container" meant that a sample was drawn from the contained population.

Preference trials

The preference trials aimed at assuring the apes' constant preference for one of the two single-item types and were conducted prior to each of the test sessions. In each trial the experimenter placed one pellet piece and one carrot piece on the sliding table close to the Plexiglas panel, directly in front of the holes. The side on which the pellet piece was positioned was counterbalanced. Apes indicated their choice with their finger and immediately received the selected food item as reinforcement. Subsequently, the test trials of the respective condition were conducted. The criterion for an ape to be included in the analysis was choosing the pellet piece in at least 75% of the trials.

Test trials

All apes participated in three conditions. To control for a possible effect of order, 15 subjects were tested in the first order of conditions

Appendix A: Eckert, Rakoczy, & Call (2017)
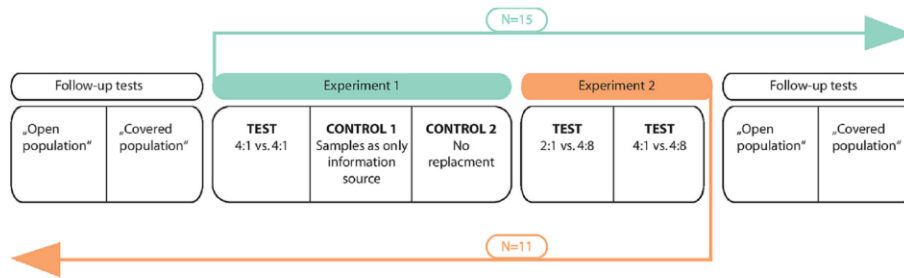
FIGURE 2 Schematic visualization of study design. Fifteen subjects were tested in the first order of conditions, starting with Experiment 1 through to Experiment 2. Eleven subjects were tested in the second order, starting with Experiment 2 through to Experiment 1. After completion of both experiments, subjects proceeded to the follow-up tests

(Test condition—Control 1—Control 2). The remaining eleven subjects were tested in the reverse order of conditions (Control 2—Control 1-Test condition). When we decided to split up subjects in the two groups of orders, all gorillas had already been tested in the first order of conditions. Thus, order was counterbalanced across subjects for all species except for the four gorillas, which were all tested in the original order of conditions (see Table 1 for information about the order of conditions each subject experienced). In all conditions, the populations consisted of 30 items each: Population A was composed of 24 pellet pieces and 6 carrot pieces; population B was composed of 6 pellet pieces and 24 carrot pieces. (These ratios were chosen because Rakoczy et al. (2014) showed that apes can reliably discriminate between multiples of the ratio 4:1 vs. 1:4. To not exceed the upper limit of caloric intake recommended for the apes, we had to limit the absolute number of food items to a certain extent. This limitation also

reduced the risk of satiation and thereby helped to keep up apes' motivation over the course of trials).

Test 4:1 versus 1:4 In this condition the samples reflected the distribution of the populations one-to-one. More specifically, the sample apparently drawn from population A (24 pellets: 6 carrots) consisted of 4 pellet and 1 carrot pieces, and the sample apparently drawn from population B (6 pellets: 24 carrots) consisted of 1 pellet and 4 carrot pieces. Before a trial started, the experimenter fixed the screen on the table to prevent the subject from watching preparations. Subsequently, she positioned the pre-prepared multi-item samples in the small compartments underneath the table. The two containers holding the populations were placed next to each other in the center of the table and the two opaque occluders were positioned over them.

The trial started when the screen was removed from the sliding table unblocking the view over the table for the subject. The experimenter simultaneously removed the two occluders from the containers, and subsequently showed each population to the ape by lifting the container, tilting it forward and shaking it slightly. After the subject had seen both populations, the experimenter repositioned the occluders over the containers and put the screen back into the metal brackets. Then she shuffled both containers. Hence, subjects knew the two available populations, but did not know which population was which. Revealing the populations at the beginning of each trial ensured that apes were aware of both containers holding a relatively high number of food items (higher than the number of items subsequently drawn). During the shuffling process, the experimenter reached into the two compartments underneath the table, retrieved the hidden samples and put them into her fists to make sure that the subject did not see them there. After removing the screen again, the experimenter pretended to draw simultaneously out of each population by inserting her fists into the two covered containers and moving them around while looking upwards (maintaining a pretence of random drawing). While the subject was watching, she simultaneously removed both hands out of the containers and presented the samples on the palms of her hands close to the Plexiglas panel saying "look!". After the ape had seen
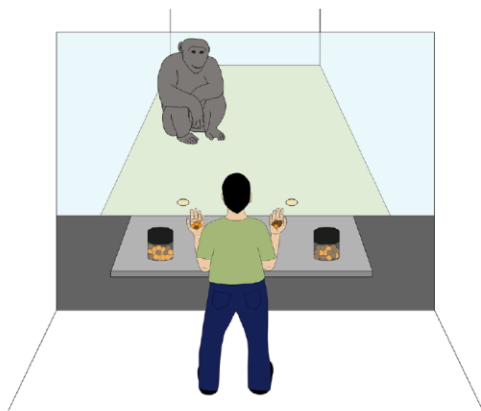
FIGURE 3 Experimental setup. The ape observed the experimenter drawing two representative multi-item samples out of two covered populations. Subsequently, the subject was given the choice between the two populations and received the content of the chosen container

both samples, the experimenter let them fall back into the containers. Subsequently, the experimenter closed her eyes to minimize unintended cueing and pushed the sliding table slightly forward so that each container, covered by an occluder, was positioned directly in front of one of the holes. By inserting a finger into one of the holes, the ape could indicate her choice, which was coded live by the experimenter after she had opened her eyes again. In cases where the subject pointed toward both containers, the sliding table was pulled backwards with the words "just one," and then pushed forward again, giving the ape a new choice between the populations. After the ape had made her decision, the occluder of the chosen container was removed, revealing the selected population. Finally, the subject received the chosen population (see Figure 1 for an illustration of the procedure).

Control 1: samples as only source of information

To investigate whether apes were able to infer from the samples alone which distribution the populations most likely had, we carried out Control 1, in which the subjects did not see the available populations prior to the sampling process. The procedure of Control 1 was the same as in the Test condition, with the following exception: In the beginning of the trials, the experimenter did not remove the two occluders from the containers, preventing the apes from seeing the two available populations. Instead, she shook the containers with the occluders consecutively, making sure that the apes were aware of something being in the containers, but leaving them in uncertainty about the exact content (see Figure 1 for an illustration of the procedure).

Control 2: no replacement of samples

One alternative explanation for subjects succeeding in the Test condition as well as in Control 1 could be that apes did not make inferences about the drawing process and the populations as a whole, but based their choices on the side where the "more attractive" sample was inserted. More specifically, apes could have tracked their preferred sample and chosen the population in which this sample was dropped in. To rule that out, we conducted Control 2, in which the samples were not re-inserted into the populations. The procedure was the same as in the Test condition, but instead of letting the samples fall back into the containers, the experimenter threw them away in a bucket next to the table. Thus, the apes were prevented from basing their choice on the side where the "more attractive" sample was inserted and could instead use the samples only as a hint for the composition of the populations (see Figure 1 for an illustration of the procedure).

Follow-up tests

A pre-requisite for the correct interpretation of results was that apes recognized and had a preference for the population containing a higher proportion of pellet pieces. Therefore, we conducted two follow-up tests. Each of them was tested within a single session consisting of four trials. Note that the follow-up tests were the last conditions subjects underwent in this study, that is, individuals that underwent Experiment 1 first, were tested in the follow-up tests after completion of Experiment 2. Subjects that were tested in Experiment 2 first, received the follow-up tests after completion of Experiment 1 (see Figure 2). This was to ensure that none of the subjects had any prior experience regarding the populations before starting the test.

"Open population"-test

In the "open population" test, apes were presented with the same populations as during test conditions (A 24:6; B 6:24). For each trial, populations were placed in transparent containers standing next to each other in the center of the sliding table. The experimenter shook both containers successively and tilted them forward to give a full view of the available populations. Once the ape had seen both populations, the experimenter positioned the containers on the edge of the sliding table, each in front of one of the holes. Subsequently, she pushed the table forward and the ape could indicate her choice by pointing through one of the holes and received the content of the chosen container. The criterion for an ape to be included in the analysis was choosing the population containing more pellets in at least 75% of trials.

"Covered population"-test

The procedure of the "covered population" test was the same as in the "open population" test, except the fact that the experimenter pulled opaque occluders on the containers after the subject had seen the content. Thus, when making a choice, the ape was prevented from seeing the two populations; instead she had to memorize the position of her preferred population for a few seconds. This second follow-up test with covered containers was conducted to test for the possibility that some apes might not have been able to choose the correct container throughout the test trials due to the fact that it was not visible when the choice had to be made. Subjects were considered successful when they chose the pellet-population in at least 75% of trials. Based on previous studies that have shown that apes can solve quantity discrimination tasks that require encoding and mental comparison of quantities [e.g., Beran, Beran, Harris, & Washburn, 2005; Call, 2000], we expected that apes would be able to cope with the type of stimuli occlusion involved in this test.

2.1.4 | Coding and data analysis

The apes' choice was coded live by the experimenter. A second blind observer coded 25% of the trials from video. Both raters were in excellent agreement ($K = 0.95$, $N = 168$). Data of five subjects (one bonobo, two chimpanzees, and two gorillas, see Table S1 for individual data) had to be excluded because those individuals did not reach criterion in the follow-up tests (see Follow-up tests section). No ape had to be excluded on the basis of the preference trials. Data of all conditions were analyzed separately using R (R Core Team, 2014). Subjects' choices were the dependent

measure and were defined as "correct" if the chosen container contained the population with the more favorable ratio of pellets to carrots (24:6). The apes' overall performance (percent correct across trials) was tested against chance level using a two-tailed one-sample t-test (R function t-test). The effect sizes were obtained applying the package "lsr" (Navarro, 2015). In addition, we tested apes' first trial performance against chance level using an exact binomial test (R function binom.test) to detect potential learning effects. In order to test whether performance differed between species we used a one-way ANOVA (R function aov). This was justified as residuals were normally distributed and homogenous as verified by visual inspection of residuals plotted against fitted values and qqplot. For Tukey's post hoc test we used the R function TukeyHSD.

2.2 | RESULTS AND DISCUSSION

2.2.1 | Test 4:1 versus 1:4

Apes as a group chose the more favorable population on average in 72% of trials (see Figure 4 and supplementary material Table S1 for individual data), significantly more often than predicted by chance (t (20) = 6.12, $p < 0.001$, 95%CI [0.64, 0.79], N = 21; Cohen's d = 1.34). This pattern was also visible in the first trial performance (Mean = 71 %; Binomial test: $p = 0.04$, N = 21; Cohen's g = 0.43). Hence, the apes' performance seems to reflect an intuitive capacity rather than a learning effect. We detected no difference between species (ANOVA: F (3, 17) = 0.2, df = 3, $p = 0.895$). These results suggest that all tested species of great apes were able to intuitively use the information provided by the samples to receive the preferred population, therefore giving a first hint toward apes being able to reason from samples to populations.

2.2.2 | Control 1: samples as only source of information

Apes as a group chose the more favorable population on average in 69% of trials (see Figure 4 and supplementary material Table S1 for individual data), which is significantly above chance level (t (20) = 5.20, $p < 0.001$, 95%CI [0.62, 0.77], N = 21; Cohen's d = 1.13). However, this
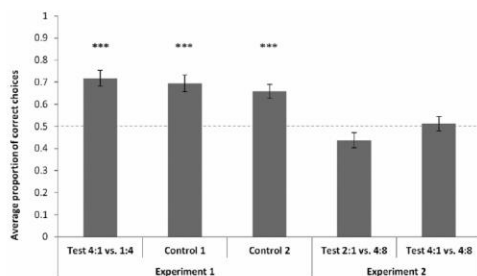
FIGURE 4 Mean proportion of trials (with standard errors) in which subjects chose the more favorable population. Dashed depicts chance level (50%)

pattern was not found considering only the performance in the first trial (Mean = 52%; Binomial test: $p = 0.5$, N = 21). This is perhaps best explained by insecurity about the available populations. Control 1 was the only condition in which subjects did not know the two possible answers (i.e., the two available populations) before making their decision. Hence, in the very first trial they could not be sure whether both populations were of the same size or whether, for example, the population associated with the "worse" sample contained many more items than the population from which the "better" sample was drawn. Potentially, apes had to experience during the first trial that, even though they had not seen the containers' content, there were two different populations of food items with the same absolute quantity. The first trial data suggest that subjects did not necessarily expect the populations to be the same as in other conditions, making it unlikely that subjects had learned and remembered the composition of the populations during the previous session(s). We detected no difference between species (ANOVA: F (3, 17) = 0.99, df = 3, $p = 0.421$). In sum, these results show that the information provided by the samples was sufficient for the apes to infer about the distribution within the populations.

2.2.3 | Control 2: no replacement of samples

Apes as a group chose the more favorable population on average in 66 % of trials (see Figure 4 supplementary material Table S1 for individual data), which is significantly more often than expected by chance (t (20) = 4.97, $p < 0.001$, 95%CI [0.59, 0.73], N = 21; Cohen's d = 1.08). This pattern was also reflected in the first trial performance (Mean = 76%; Binomial test: $p = 0.01$, N = 21; Cohen's g = 0.52) and thus cannot be due to learning. In this condition we detected differences between species (ANOVA: F (3, 17) = 4.88, df = 3, $p = 0.01$, R^2 = 0.46). Tukey multiple comparison of means revealed that bonobos performed significantly worse than gorillas (Mean bonobos = 53%, N = 5; Mean gorillas = 88%, N = 2, $p = 0.015$). However, considering the fact that we could only include the data of two gorillas (compared to five bonobos) in the final analysis, it is questionable whether this result truly reflects differences between species, or rather random variation or individual differences between subjects. The findings of Control 2 rule out the possibility that the apes solved the task by means of a simple heuristic: "choose the container where the more attractive sample was inserted." Instead, apes seem to have considered the drawing process and inferred about the population as a whole.

In sum, the results of Experiment 1 show that all tested species of great apes were able to use information provided by multi-item samples to track their preferred populations, and they did so even when they did not know the composition of the populations beforehand (Control 1) and when samples were not replaced after drawing (Control 2). These findings suggest that great apes might engage in intuitive statistical inferences from samples to populations in a comparable way human infants do (Xu & Garcia 2008; Denison et al., 2013). However, an alternative explanation for these results could be that apes simply associated the preferable sample (i.e., the sample containing absolutely more pellets), with

the container that it was drawn from. To address this alternative explanation, we tested subjects in Experiment 2 with samples in which absolute and relative frequencies of pellets were disentangled.

3 | EXPERIMENT 2: DO APES TAKE INTO ACCOUNT RELATIVE, RATHER THAN ABSOLUTE FREQUENCIES?

Although results of Experiment 1 tentatively suggest that apes were able to reason from multi-item samples to populations, it is an open question to what extent the subjects relied on absolute quantities rather than on proportions to solve the task. More specifically, in all conditions of Experiment 1, absolute and relative frequencies were confounded within the samples, that is, the sample which contained the higher proportion of preferred food items than the alternative (4:1 vs. 1:4), also contained the higher absolute quantity of preferred food items (4 vs. 1). Thus, Experiment 1 alone cannot tease apart whether apes truly compared the proportion of pellets to carrots in both samples (4:1 vs. 1:4), or if they based their choice on the absolute amount of pellets (4 vs. 1) and used the heuristic: "choose the container where more pellets were drawn from." To address this question we tested apes in Experiment 2 in two further conditions. In both of them, absolute and relative frequencies within the samples were arranged in such a way that apes could not perform above chance level if they focused on absolute numbers only (see Figure 1 for an illustration of the Test conditions).

3.1 | Methods

3.1.1 | Subjects

The same 26 individuals as in Experiment 1 participated in this experiment. One additional chimpanzee was tested but excluded from data analysis as he did not complete all sessions due to a lack of motivation.

3.1.2 | Materials

We used the same materials as in Experiment 1 (see Figure 3 for an illustration of the experimental setup).

3.1.3 | Design and procedure

The general procedure was the same as in Experiment 1. To tease apart whether apes truly compared the proportion of preferred to neutral food items in both samples, or if they based their choice on the absolute amount of preferred food, we tested apes in two conditions with varying sample composition. Again, each condition consisted of 12 test trials, divided into three sessions. Prior to the test trials, two preference trials with single pellet and carrot pieces were carried out. Thus, each session consisted of two preference trials and four test trials.

Familiarization

Each subject that had not experienced Experiment 1 before received one session with six trials of familiarization. The procedure of the familiarization phase was exactly as described for Experiment 1.

Preference trials

The procedure of the preference trials was the same as in Experiment 1.

Test trials

All apes participated in two Test conditions. To control for a possible effect of order, 15 subjects were tested in the first order of conditions, starting with the 2:1 versus 4:8 test, through to the 4:1 versus 4:8 test. The remaining eleven subjects were tested in the reverse order of conditions (see Table 1 for information about the order of conditions each subject experienced). Again, in all conditions the populations consisted of 30 items each: Population A was composed of 24 pellet pieces and 6 carrot pieces; population B was composed of 6 pellet pieces and 24 carrot pieces

Test 2:1 versus 4:8 The procedure was the same as described for the Test condition of Experiment 1. However, the composition of the samples was varied in such a way that choosing the container from which the sample with the higher absolute number of pellets was drawn, resulted in receiving the less attractive population. In particular, the sample apparently drawn from population A (24 pellets: 6 carrots) consisted of 2 pellet and 1 carrot pieces, and the sample apparently drawn from population B (6 pellets: 24 carrots) consisted of 4 pellet and 8 carrot pieces. Thus, even though sample B contained double the amount of pellets compared to sample A, the proportion of pellets to carrots was more favorable in sample A. If apes' choice was based on absolute quantities, we expected them to choose the "wrong" container more often than the "correct" one. If they, however, took into account the proportion of pellets to carrots, we expected them to choose the "correct" container more often than the foil (see Figure 1 for an illustration of the procedure).

6.4.2 | Test 4:1 versus 4:8

Again, the procedure was the same as described for the Test condition of Experiment 1. However, here the composition of the samples was varied in a way that both samples contained the same absolute number of pellets. More specifically, the sample apparently drawn from population A (24 pellets: 6 carrots) consisted of 4 pellet and 1 carrot pieces, and the sample apparently drawn from population B (6 pellets: 24 carrots) consisted of 4 pellet and 8 carrot pieces. Assuming that apes based their choice on absolute quantities only, we expected them to choose both containers at similar rates, as the absolute number of

pellets did not provide any conclusive information. If they instead reasoned about the proportion of pellets to carrots, we predicted that they chose the correct container more often than expected by chance (see Figure 1 for an illustration of the procedure).

Follow-up tests

Those individuals that underwent Experiment 2 after Experiment 1 received the two follow-up tests. The procedure was exactly the same as described for Experiment 1.

3.1.4 | Coding and data analysis

The apes' choice was coded live by the experimenter. A second blind observer coded 25% of the trials from video. Both raters were in excellent agreement ($K = 0.95$, $N = 120$). Data of five subjects (one bonobo, two chimpanzees, and two gorillas, see Table S1 for individual data) had to be excluded because those individuals did not reach criterion in the follow-up tests. No further ape had to be excluded on the basis of the preference trials. Data analysis was the same as described for Experiment 1.

3.2 | Results and discussion

3.2.1 | Test 2:1 versus 4:8

Apes as a group chose the more favorable population on average in 44% of trials (see Figure 4 and supplementary material Table S1 for individual data). Though this pattern is not different from what was expected by chance (t (20) = −1.84, $p = 0.08$, 95%CI [0.36, 0.51], $N = 21$), it indicates a (non-significant) trend such that apes tended to choose the less favorable population more often than the more favorable one. We detected no differences between species (ANOVA: F (3, 17) = 1.66, df = 3, $p = 0.213$). This pattern was also reflected in the first trial performance (Mean = 47%; Binomial test: $p = 1$, $N = 21$). Hence, all tested species of great apes were unable to extrapolate from samples to populations, when the absolute number of preferred food-items was misleading. Instead, they tended to choose the population where the sample with the higher amount of preferred food-items was drawn from. This finding gives a first hint that the strategy applied by the apes might have been a comparison of absolute numbers between samples, rather than an extrapolation of proportions.

3.2.2 | Test 4:1 versus 4:8

Apes as a group chose the more favorable population on average in 51% of trials (see Figure 4 and supplementary material Table S1 for individual data), which is not different from chance level (t (20) = 0.37, $p = 0.715$, 95%CI [0.44, 0.58], $N = 21$). We detected no differences between species (ANOVA: F (3, 17) = 1.35, df = 3, $p = 0.292$). The same pattern was found considering only the performance in the first trial (Mean = 43%; Binomial test: $p = 0.664$, $N = 21$). This implies that apes failed to use the information provided by the samples to reason about

the populations and strengthens the theory that apes might have relied on absolute, rather than relative frequencies.

4 | GENERAL DISCUSSION

In Experiment 1, we investigated whether great apes are able to reason from multi-item samples to populations of food items. Results showed that great apes did extrapolate from samples to populations, irrespective of whether they knew the composition of the available populations beforehand or not (Control 1) and if samples were replaced after drawing or not (Control 2). The results of Control 2 are especially revealing, as they rule out the possibility of a simple heuristic: "choose the container where the more attractive sample was inserted". Instead, apes seem to have considered the drawing process and inferred about the population as a whole from the first trial onwards. This implies that apes seem to possess similar kinds of capacities as found in human infants (Xu & Garcia, 2008; Denison et al., 2013). In fact, our findings even go one step further than those of the two existing studies that tested infants' ability to reason from sample to population: While the apes in our study drew inferences from samples to populations in an active choice paradigm, the human infants in the above mentioned studies were only tested using the VOE looking-time paradigm. There is some evidence that findings of studies using the VOE looking time paradigm dissociate from findings of studies using active choice measures [e.g., Ahmed & Ruffman, 1998; Shinskey & Munakata, 2005; Charles & Rivera, 2009]. This is probably due to the fact that a subject that is able to perceive something is not necessarily able to act accordingly. As it is currently unknown whether human infants would succeed in an active choice paradigm testing for their capacities to reason from sample to population, we conclude that great apes' intuitive statistical abilities in this regard seem to be at least at a comparable level as those of young human infants. However, based on Experiment 1 alone it is impossible to rule out that apes used alternative strategies based on the absolute number of preferred food items. The aim of Experiment 2, therefore, was to investigate whether great apes can successfully reason from samples to populations when prevented from relying on absolute quantities. Apes performed at chance level both when the sample drawn from the more favorable population contained less preferred food items than the sample drawn from the less favorable population, and when both samples contained the same number of preferred food items. Thus, apes did not rely on inferences from samples to populations in this experiment. There are at least two interpretations for these findings.

One interpretation is that apes' failure in Experiment 2 reflects true limitations of their cognitive competences. The most obvious difference between Experiment 1 and 2 is that only in the latter subjects could not rely on absolute numbers of preferred food items. Hence, one could conclude that apes are able to reason and draw inferences about absolute, but not relative frequencies. Assuming that apes simply compared the absolute quantity of pellets in both samples and chose the population from which more pellets were drawn, we expected the following pattern of results: When the number of pellets

in the samples was inconclusive (because it was the same in both samples), apes should have chosen randomly between both populations. When the number of pellets was misleading, that is, higher in the sample drawn from the non-preferred population, apes should have chosen the "wrong" population more often. While apes indeed chose randomly between populations when the number of pellets was the same in both samples, they also did so when the number of pellets was misleading. Yet, it should be noted that even though there was no significant effect in this condition (misleading number of pellets in both samples), apes nevertheless revealed a non-significant tendency to choose the more favorable population less often than the more favorable one. Consequently, it cannot be ruled out that apes mainly relied on absolute quantities in this experiment.

This opens up an alternative explanation for the apes' success in Experiment 1: Subjects might have not drawn any inference from sample to population, but instead simply associated the more favorable sample (i.e., the one containing absolutely more preferred items than the other) with the container it was drawn from, since it was temporally and spatially most closely associated with that container. In other words, apes might have followed a heuristic like "chose the container where you saw something good (i.e., more pellets) coming from." Future studies need to determine whether subjects truly relied on associating containers with "better" and "worse," or if they in fact perceived the samples as a representation of populations. One possible way to disentangle the two explanations would entail presenting apes with two opaque containers filled with two populations of food items (similar to the current study). Crucially, the experimenter would already have the samples in her hands (i.e., pellets and carrots in 4:1 distribution in one hand, 1:4 in the other). She would then show the contents of her hands to the ape, insert her hands into the containers and remove them again, showing the same items as before. Subsequently, she would discard the samples and give the apes the choice between the two containers. If apes merely associated the two containers with "good" or "bad" according to the distribution they had seen on each side, we would expect them to choose the side where the sample with absolutely more pellets was shown. In contrast, if they recognized a randomly drawn sample as representation of the population, they should pick both containers equally often since no drawing took place, and therefore, no inference can be made.

Recall that Rakoczy et al. (2014) showed that great apes did take proportions into account when reasoning the other way around, that is from populations to samples, ruling out that subjects used a simple association mechanism to solve the task. If our results reflected true limitations in apes' cognitive competences, they would, therefore, suggest that nonhuman primates' statistical abilities could be unidirectional. This would question whether apes have a true understanding of drawing processes and the relation between populations and samples.

A different interpretation for the negative findings of Experiment 2 is that they may merely reflect performance limitations imposed by the task's cognitive demands, which may have masked apes' true competence. One of these task demands could be the memory component required by our procedure. At the exact moment when apes were asked to make a choice, the information necessary to do so (i.e., the samples) was not available anymore. Instead, apes had to memorize this information for a few seconds and recall it to choose between the two populations. Note that this was not the case in Rakoczy et al. (2014), where subjects were still able to see the populations during their choice. Even though it may seem trivial to remember information for a few seconds, results of the follow-up test with covered populations showed that this was indeed a crucial factor for some of the subjects: Four of the 26 subjects were not able to choose the more attractive population when it was covered while the decision was made, even though they showed a clear preference for that population during the preference test with open populations. Furthermore, other studies have shown the importance of working memory in different problem solving tasks. For instance, in Seed, Seddon, Greene, and Call (2012) four chimpanzees solved a tool-use task requiring causal inferences when the time-span over which information had to be memorized was minimized. By contrast, in a related previous study (Povinelli, 2000) that involved a higher working memory load, all chimpanzees failed to do so. Although working memory demands, potentially in combination with lack of attention, may have influenced the apes' performance to a certain extent, working memory alone cannot fully explain the fact that apes were not able to use proportional information in this experiment. Recall that those subjects who had difficulties remembering the populations' position were excluded from the analysis and did therefore not bias the results in a negative way. Moreover, Experiment 1 also required a memory component, and still subjects succeeded.

Another factor that could have made this task more difficult as compared to Rakoczy et al. (2014) is the type of inferences required. Retrospective inferences seem to be harder than prospective ones (Völter & Call, 2017). This means that going from samples back to populations (retrospective) may be more demanding than going from populations forward to samples (prospective). The majority of knowledge that we have about the origin and development of intuitive statistics derives from the extensive study of pre-verbal infants. In the last decades, numerous such studies have tested infants both for their abilities in reasoning from populations to samples as well as from samples to populations. As mentioned above, to our knowledge there is no study testing pre-verbal infants for their ability to reason from samples to populations in an active choice measure. This type of methodology was, so far, only used in studies investigating infants' capacity to reason from population to sample (Feigenson et al., 2002; Denison & Xu, 2010b; Denison & Xu, 2014). In these studies, infants were allowed to choose between the covered samples of two populations of preferred and non-preferred items in different ratios. Control conditions disentangled absolute and relative frequencies with the result that infants indeed used proportional information, not a comparison of absolute quantities, to retrieve their preferred item. The two existing studies

investigating the reverse ability, that is, reasoning from samples to populations (Xu & Garcia, 2008; Denison et al., 2013), both used a VOE looking-time paradigm, a methodology that is less comparable with the methodology applied for great apes. Moreover, in both above-mentioned studies probability was confounded with quantity, and no control condition tested for the fact that infants could have used the shortcut of focusing on absolute quantities only. As a consequence, it remains unclear whether reasoning from samples to populations represents a cognitively more challenging task than the other way around. It would be of great interest to fill that gap of knowledge by applying an active choice paradigm to investigate pre-verbal infants' ability to reason from samples to populations, including a control condition for absolute versus relative information.

A third task demand that may have masked apes' true competence in Experiment 2 is the poorer discriminability of the samples as compared to the samples used in Experiment 1. As an index for discriminability we calculated the ratio of ratios (hereafter: ROR) of the two samples for each of the conditions in the following way [following Drucker, Rossa, & Brannon, 2016]:

$$\frac{\text{Ratio of pellets to carrots in the sample drawn from the preferred population}}{\text{Ratio of pellets to carrots in the sample drawn from the non-preferred population}}$$

In all conditions of Experiment 1, the ROR was (4/1)/(1/4) = 16 (in Rakoczy et al., 2014 the ROR was ≥16 in all conditions). In Experiment 2, the ROR was (2/1)/(4/8) = 4 in the 2:1 versus 4:8 test, and (4/1)/(4/8) = 8 in the 4:1 versus 4:8 test. Thus, in both conditions of Experiment 2, the ROR was less than or equal to half the one used in Experiment 1. This discrepancy was caused by our methodological constraints that prevented us from using larger RORs. More specifically, a larger ROR would have required larger samples and thus larger populations. As the apes received the "whole" chosen population as reinforcement we had to minimize the number of food items within the populations for the purpose of not exceeding their allowed daily caloric intake. Moreover, given that the food items were kept in the experimenter's fist, larger samples would have required a different sampling method than the one applied here. As a consequence, in this study it was not possible to disentangle absolute and relative information with the same ROR as in Experiment 1. Recent research suggests that indeed the magnitude of difference between two proportions is crucial for non-human primates to discriminate probabilities. Hanus & Call (2014) presented chimpanzees with two trays, each of them with a different ratio of hidden food items to potential hiding locations and therefore a different likelihood of finding food. This study revealed that subjects' performance was influenced by the relative difference between the two probabilities as soon as a certain threshold thereof was reached. Moreover, the apes relied on the ratio between probabilities, even in conditions where one tray depicted an absolute safe option—a probability of finding food of 100%. This study emphasizes the importance of the magnitude of difference

between the two ratios to be discriminated, rather than the magnitude of difference within the single ratios.

With regard to the present study this means the following: Although the quantities within one sample were presumably easy to discriminate [for reviews about quantity discrimination see example, Feigenson, Dehaene, & Spelke, 2004; Nieder, 2005], it was probably the ratio between the ratios of both samples that influenced the decision of the apes and it could well be that the present RORs were simply below the threshold for discriminating two ratios and thus failed to constitute notable differences. In a study using a touch screen setup (Drucker et al., 2016) rhesus macaques (*Macaca mulatta*) were presented with arrays containing different ratios of positive to negative stimuli. The monkeys learned to choose those arrays with the greater ratio of positive to negative stimuli and were able to generalize to novel ratios. Similarly, as in the previously mentioned study with chimpanzees (Hanus & Call, 2014), the performance was directly influenced by the magnitude of difference between the two ratios to be discriminated. Interestingly, just as human infants (McCrink & Wynn, 2007), the two macaques tested were able to discriminate a ROR of 2, which is much lower than those used in our experiments. However, given the fact that those subjects received extensive training in such discrimination tasks before the actual test, it remains unclear to which extent those methods are comparable to the ones used here with apes.

5 | CONCLUSIONS

The aim of the current study was to investigate whether apes can use samples of items to infer the composition of the population from where the samples came from. While apes performed competently when the samples from the more favorable population were more attractive than the samples from the less favorable population not only in terms of relative but also in terms of absolute frequencies of preferred over non-preferred food items, they failed to do so when absolute and relative frequencies were disentangled. The present study, therefore, cannot determine whether non-human primates engage in intuitive statistical inferences from randomly drawn samples to populations in a comparable way human infants have recently been found to do (Denison et al., 2013; Xu & Garcia, 2008). It is an open question for future research whether these limitations in apes' performance reflect true limits of cognitive competence or merely performance limitations due to accessory task demands.

CONFLICT OF INTEREST

None.

ORCID

Johanna Eckert http://orcid.org/0000-0001-8161-0828

REFERENCES

Ahmed, A., & Ruffman, T. (1998). Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a nonsearch task? *Developmental Psychology, 34*(3), 441–453.

Barnard, A. M., Hughes, K. D., & Gerhardt, R. R., DiVincenti, L., Bovee, J. M., & Cantlon, J. F. (2013). Inherently analog quantity representations in olive baboons *(Papio anubis). Frontiers in Psychology, 4,* 253.

Benson-Amram, S., Heinen, V. K., Dryer, S. L., & Holekamp, K. E. (2011). Numerical assessment and individual call discrimination by wild spotted hyaenas, *Crocuta Crotua. Animal Behaviour, 82,* 743–752.

Beran, M. J., Beran, M. M., Harris, E. H., & Washburn, D. A. (2005). Ordinal judgments and summation of nonvisible sets of food items by two chimpanzees *(Pan troglodytes) and a rhesus macaque (Macaca mulatta). Journal of Experimental Psychology: Animal Behavior Processes, 31,* 351–362.

Beran, M. J., Evans, T. A., Leighty, K. A., Harris, E. H., & Rice, D. (2008). Summation and quantity judgments of sequentially presented sets by capuchin monkeys *(Cebus apella). American Journal of Primatology, 70*(2), 191–194.

Beran, M. J., McIntyre, J. M., Garland, A., & Evans, T. A. (2013). What counts for 'counting'? Chimpanzees, Pan troglodytes, respond appropriately to relevant and irrelevant information in a quantity judgment task. *Animal Behaviour, 85*(5), 987–993.

Beran, M. J., & Parrish, A. E. (2016). Capuchin monkeys *(Cebus apella)* treat small and large numbers of items similarly during a relative quantity judgment task. *Psychonomic Bulletin and Review, 23*(4), 1206–1213.

Boysen, S. T., & Berntson, G. G. (1989). Numerical competence in a chimpanzee *(Pan troglodytes). Journal of Comparative Psychology, 103*(1), 23–31.

Call, J. (2000). Estimating and operating on discrete quantities in orangutans *(Pongo pygmaeus). Journal of Comparative Psychology, 114*(2), 136–147.

Charles, E. P., & Rivera, S. M. (2009). Object permanence and method of disappearance: Looking measures further contradict reaching measures. *Developmental Science, 12*(6), 991–1006.

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.

Dacke, M., & Srinivasan, M. V. (2008). Evidence for counting in insects. *Animal Cognition, 11*(4), 683–689.

Davis, H. (1984). Discrimination of the number three by a raccoon. *Animal Learning & Behavior, 12,* 409–413.

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.

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

Denison, S., & Xu, F. (2010a). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science, 34*(5), 885–908.

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.

Denison, S., & Xu, F. (2012). Probabilistic inference in human infants. *Advances in Child Development and Behavior, 43,* 27–58.

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition, 130*(3), 335–347.

Drucker, C. B., Rossa, M. A., & Brannon, E. M. (2016). Comparison of discrete ratios by rhesus macaques *(Macaca mulatta). Animal Cognition, 19*(1), 75–89.

Farnsworth, G. L., & Smolinski, J. L. (2006). Numerical discrimination by wild northern mockingbirds. *Condor, 108*(4), 953–957.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science, 13*(2), 150–156.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.

Geary, D. C., Berch, B. B., & Mann Koepke, K. (2015). The evolution of number systems. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Evolutionary origins and early development of number processing* (pp. 335–353). San Diego, CA: Academic Press.

Hamilton I. M., (2010). Foraging theory. In D. F. Westneat, & C. W. Fox, (Eds.), *Evolutionary behavioral ecology* (pp. 177–193). New York: Oxford University Press.

Hanus, D., & Call, J. (2007). Discrete quantity judgments in the great apes *(Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus)*: The effect of presenting whole sets versus item-by-item. *Journal of Comparative Psychology, 121*(3), 241–249.

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

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960.

Hunt, S., Low, J., & Burns, K. C. (2008). Adaptive numerical competency in a food-hoarding songbird. *Proceedings of the Royal Society B: Biological Sciences, 275*(1649), 2373–2379.

McComb, K., Packer, C., & Pusey, A. (1994). Roaring and numerical assessment in contests between groups of female lions, *Panthera leo. Animal Behaviour, 47*(2), 379–387.

McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science, 18,* 740–745. https://doi.org/10.1111/j.1467-9280.2007.01969.x

Navarro D. J. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.5).* Adelaide, Australia: University of Adelaide.

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience, 6*(3), 177–190.

Perdue, B. M., Talbot, C. F., Stone, A. M., & Beran, M. J. (2012). Putting the elephant back in the herd: Relative quantity judgments match those of other species. *Animal Cognition, 15,* 955–961.

Piaget J., & Inhelder B. (1975). *The origin of the idea of chance in children.* New York: Norton, (p. 251).

Pisa, P., & Agrillo, C. (2009). Quantity discrimination in felines: A preliminary investigation of the domestic cat *(Felis silvestris catus). Journal of Ethology, 27*(2), 289–293.

Potrich, D., Sovrano, V. A., Stancher, G., & Vallortigara, G. (2015). Quantity discrimination by zebrafish *(Danio rerio). Journal of Comparative Psychology, 129*(4), 388–393.

Povinelli, D. J. (2000). *Folk physics for apes: The chimpanzee's theory of how the world works.* Oxford University Press, Oxford: New York.

R Core Team, 2014. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rakoczy, H., Clüver, A., & Saucke, L., Stoffregen, N., Grabener, A., Migura, J., & Call, J. (2014). Apes are intuitive statisticians. *Cognition, 131*(1), 60–68.

Reznikova, Z., & Ryabko, B. (2011). Numerical competence in animals, with an insight from ants. *Behaviour, 148*(4), 405–434.

Rugani, R., Cavazzana, A., Vallortigara, G., & Regolin, L. (2013). One, two, three, four, or is there something more? Numerical discrimination in day-old domestic chicks. *Animal Cognition, 16*(4), 557–564.

Seed, A., Seddon, E., Greene, B., & Call, J. (2012). Chimpanzee 'folk physics': bringing failures into focus. *Philosophical Transactions of the Royal Society B Biological Sciences, 367*, 2743–2752. https://doi.org/10.1098/rstb.2012.0222

Shinskey, J. L., & Munakata, Y. (2005). Familiarity breeds searching — Infants reverse their novelty preferences when reaching for hidden objects. *Psychological Science, 16*(8), 596–600.

Skyrms B. (1975). *Choice and chance: An introduction to inductive logic.* Encino, CA: Dickenson Pub, (p. 220).

Tecwyn, E. C., Denison, S., Messer, E. J., & Buchsbaum, D. (2016). Intuitive probabilistic inference in capuchin monkeys. *Animal Cognition, 130*(3), 335–347.

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 USA, 104*(48), 19156–19159.

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.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*(7), 309–318.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty—Heuristics and biases. *Science, 185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458.

Vonk, J., & Beran, M. J. (2012). Bears "count" too: Quantity estimation and comparison in black bears, *Ursus americanus. Animal Behaviour, 84*, 231–238.

Völter, C. J., & Call, J. (2017). Causal and inferential reasoning in animals. In G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), APA handbook of comparative psychology Vol 2: Perception, learning, and cognition (pp. 643–671). US: American Psychological Association.

Ward, C., & Smuts, B. B. (2007). Quantity-based judgments in the domestic dog *(Canis lupus familiaris). Animal Cognition, 10*(1), 71–80.

Wilson, M. L., Britton, N. F., & Franks, N. R. (2002). Chimpanzees and the mathematics of battle. *Proceedings of the Royal Society of London B Biological Sciences, 269*, 1107–1112.

Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition, 112*, 97–104.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences USA, 105*(13), 5012–5015.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Appendix B: Eckert, Call, Hermes, Herrmann, & Rakoczy (2018)

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. doi: 10.1016/j.cognition.2018.07.004

Contents lists available at ScienceDirect

Cognition

Original Articles

Intuitive statistical inferences in chimpanzees and humans follow Weber's law

Johanna Eckert[a,b,c,*], Josep Call[a,d], Jonas Hermes[b,c], Esther Herrmann[a], Hannes Rakoczy[b,c]

[a] Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
[b] Department of Developmental Psychology, University of Goettingen, Waldweg 26, 37073 Goettingen, Germany
[c] Leibniz ScienceCampus "Primate Cognition", German Primate Center/Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Goettingen, Germany
[d] School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, UK

ABSTRACT

Humans and nonhuman great apes share a sense for intuitive statistical reasoning, making intuitive probability judgments based on proportional information. This ability is of fundamental importance, in particular for inferring general regularities from finite numbers of observations and, vice versa, for predicting the outcome of single events using prior information. To date it remains unclear which cognitive mechanism underlies and enables this capacity. The aim of the present study was to gain deeper insights into the cognitive structure of intuitive statistics by probing its signatures in chimpanzees and humans. We tested 24 sanctuary-living chimpanzees in a previously established paradigm which required them to reason from populations of food items with different ratios of preferred (peanuts) and non-preferred items (carrot pieces) to randomly drawn samples. In a series of eight test conditions, the ratio between the two ratios to be discriminated (ROR) was systematically varied ranging from 1 (same proportions in both populations) to 16 (high magnitude of difference between populations). One hundred and forty-four human adults were tested in a computerized version of the same task. The main result was that both chimpanzee and human performance varied as a function of the log(ROR) and thus followed Weber's law. This suggests that intuitive statistical reasoning relies on the same cognitive mechanism that is used for comparing absolute quantities, namely the analogue magnitude system.

1. Introduction

Statistical reasoning is of fundamental importance in human life and one of the hallmarks of human thinking: we continually generalize from sample observations and use these generalizations to predict the outcome of events and to make rational decisions under uncertainty. Nevertheless, over many decades statistical reasoning was deemed to be dependent on language and mathematical training and to remain difficult and error-prone throughout an individual's lifespan (e.g. Piaget & Inhelder, 1975; Tversky & Kahneman, 1974, 1981). More recently, however, developmental research produced evidence that even preverbal infants are capable of basic forms of intuitive statistics: For example, when confronted with two jars containing mixtures of attractive and neutral candy in different proportions, infants were able to infer which of the two was more likely to lead to a preferred candy as randomly drawn sample (Denison & Xu, 2010a, 2014). This also works in the other direction: When confronted with samples, infants could draw conclusions about the proportional composition of the associated populations (Denison, Reed, & Xu, 2013; Xu & Garcia, 2008). Even more impressively, such statistical information is integrated with knowledge from other cognitive domains from very early on: Infants seem to understand that a sample does not necessarily reflect the population's distribution, for instance when the experimenter has the intention to draw a certain type of object as well as visual access to the population (Xu & Denison, 2009), or when a mechanical constraint prevents her from drawing some of the objects (Denison, Trikutam, & Xu, 2014; Denison & Xu, 2010b). Similarly, infants can integrate complex spatio-temporal information into their statistical inferences to judge single event probabilities: When a population of objects bounced randomly in a container with one opening, infants formed expectations as to which object was most likely to exit, based on the proportional composition of the population (majority objects are more likely to exit) and each objects' spatial distance from the opening (the closer objects are, the more likely they are to exit; Teglas et al., 2011).

These findings demonstrated that neither language nor mathematical education are prerequisites for basic statistical reasoning. Instead,

* Corresponding author at: Deutscher Platz 6, 04103 Leipzig, Germany.
  E-mail address: johanna.eckert@eva.mpg.de (J. Eckert).

infants seem to be equipped with an efficient and powerful statistical inference mechanism from very early on, which presumably helps them to rapidly learn about rules and regularities of the world. From a comparative point of view, these findings raise the question of whether intuitive statistics may be part of our evolutionary heritage and therefore not necessarily uniquely human.

To shed light on this question, comparative research adapted Denison and Xu's (2010) paradigm and tested nonhuman great apes for their statistical reasoning capacities: Rakoczy et al. (2014) presented chimpanzees, gorillas, orangutans and bonobos with two transparent buckets containing mixtures of preferred and non-preferred food items (banana pellets and carrot pieces) in specific ratios. Subsequently, the experimenter drew one sample each and the subject was allowed to choose between the two covered samples. Apes were able to infer which of the two populations was more likely to lead to a pellet as a sample. Moreover, they chose systematically even when absolute and relative frequencies were disentangled, i.e. when the population with the more favorable ratio of pellets to carrots contained absolutely fewer pellets than the other one. A very recent study (Eckert, Rakoczy, Call, Herrmann, & Hanus, 2018) showed that these statistical inferences are not an isolated and automatic process; instead, just like human infants, apes did consider additional information about the experimenter (her preferences and visual access) when predicting the outcome of her draw. Hence, some great ape intuitive statistical abilities seem to be on a par with those of human infants, suggesting that they constitute an evolutionary ancient ability.

However, when great apes were presented with the reverse task, requiring inferences from sample to population, they exhibited some limitations: Eckert, Rakoczy, and Call (2017) presented apes with covered containers holding populations of preferred and non-preferred food-items. After observing multi-item samples being drawn from these populations, apes could choose between the two covered containers. Subjects were able to correctly reason from sample to population, but only in conditions, in which the proportionally favorable sample also contained absolutely more preferred food items than the other (4:1 vs. 1:4 preferred to non-preferred food items). In experiments contrasting absolute and relative frequencies of preferred food items (e.g. 2:1 vs. 4:8 preferred to non-preferred food items), apes tended to choose the population from which the sample with absolutely more preferred items was drawn (4:8), despite its unfavorable ratio.

There are at least two possible interpretations of these findings: First, they may indicate that apes relied on absolute quantity heuristics to reason from sample to population. In fact, even in the human literature there is a great deal of research suggesting that most probabilistic inferences are actually just the result of different heuristics (e.g. Davidson, 1995; De Neys & Vanderputte, 2011; Jacobs & Potenza, 1991; Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1974, 1981). Hence, controlling for absolute number heuristics is of substantial importance in the field of intuitive statistics. Nevertheless, many studies (both on human and nonhuman species) have unfortunately failed to sufficiently control for such strategies. The previously described studies on great apes, for instance, included control conditions for heuristics dealing with the absolute quantity of preferred items (which apes passed in Rakoczy et al. (2014), but failed in Eckert et al. (2017)). None of them, however, tested for the reverse strategy: a heuristic based on avoiding the population or sample containing more non-preferred food items. Hence, to be able to draw conclusions about apes' intuitive statistical abilities, there is an urgent need for studies controlling for all types of absolute quantity heuristics.

A second interpretation for Eckert et al.'s (2017) negative findings is that they merely reflect performance, rather than competence limitations. The critical conditions in that study may have been especially difficult because the magnitude of difference between samples (i.e. the ratio of the two ratios, ROR) was relatively small, and perhaps beyond the signature limits of apes' capacity: While the ROR was 16 in the successful confounded conditions (and also in all conditions of Rakoczy

and colleagues' population to sample study), it was as low as 4 in the critical condition in which absolute and relative frequencies of preferred food items were disentangled.

This, in turn, raises a much more fundamental question: What are the cognitive foundations of intuitive statistics that explain both the scope and limits of this capacity? And are these the same in humans and our closest living relatives, the chimpanzees?

Nonhuman primates (and other animals) share with humans a cognitive mechanism for basic quantitative cognition. This mechanism, the analogue magnitude system, is used for dealing with absolute numerical information: It represents number (and also other magnitudes like duration or space) by a mental magnitude that is roughly proportional to and thus a direct *analogue* of the number of individuals in the set being enumerated (see, e.g. Carey, 2009; Dehaene, 2011).[1] It thus enables subjects to estimate and compare arbitrarily large quantities, but only in an approximate way (e.g. Nieder & Dehaene, 2009). Its accuracy follows Weber's Law: Discriminability of two sets varies as a function of the ratio of the set sizes to be compared, independent of their absolute numerosity (e.g. Cantlon & Brannon, 2006, 2007). For example, if a subject can discriminate 2 from 4 objects, it is also able to discriminate 10 from 20 or 500 from 1000. This fundamental characteristic yields specific signatures that can be used to identify the involvement of this system in cognitive tasks. Numerous comparative studies have shown that many species across the animal kingdom exhibit the same signatures in accordance with Weber's law when confronted with quantity comparison tasks. For example, fish (Buckingham, Wong, & Rosenthal, 2007), birds (Ain, Giret, Grand, Kreutzer, & Bovet, 2009; Rugani, Cavazzana, Vallortigara, & Regolin, 2013), monkeys (Barnard et al., 2013; Cantlon & Brannon, 2007), and great apes (Beran, 2004; Call, 2000; Hanus & Call, 2007) all showed a ratio-dependent performance when discriminating absolute quantities, suggesting that the analogue magnitude system is an evolutionary ancient mechanism (see Beran, 2017 for a review).

Are the limits reported by Eckert et al. (2017) a first hint that apes employed their analogue magnitude system in a statistical reasoning task as well? Is the ability to reason probabilistically from population to sample and vice versa dependent on the ratio between the two proportions to be discriminated, i.e. on the ratio of ratios (ROR)? To our knowledge, no study has directly tested for this hypothesis in nonhuman primates yet; indeed, even in human adults the evidence is ambiguous. One study (O'Grady, Griffiths, & Xu, 2016) tested human adults in a computer based task that required them to reason statistically from population to sample. The authors varied the magnitude of difference between the two populations' ratios and included both trials in which absolute and relative frequencies of target stimuli were confounded and trials in which they were disentangled. Statistical analysis revealed that the effect of the ROR on humans' performance was dependent on the trial type, with much stronger effects in the confounded condition. In this condition, participants could simply compare the two (absolute) amounts of target stimuli, a capacity known to be enabled by the analogue magnitude system. Unfortunately, it remained unclear whether there was any significant ROR impact in the crucial trials controlling for absolute quantity heuristics.[2] Hence, this study is yet

---

[1] We prefer to refer to the more general "analogue magnitude system" rather than the more specific "approximate number system" since this leaves open the possibility that the system in question is not restricted to discrete numerical information in the more narrow sense, but potentially also represents continuous magnitudes such as length, duration, etc. Whether or not there is a separate cognitive system processing numerical information only is still highly debated (see e.g. Lourenco, 2015 for a review).

[2] In fact, the authors found a significant three-way-interaction between ROR, condition and age. Hence, the effect of the ROR was not only dependent on the trial type, but also on the age of participants (yet this age effect was not discussed in the paper). It is generally not meaningful or reasonable to interpret the individual effects of the components of a significant interaction (Bortz,

another example of the problematic confound of absolute and relative frequencies in intuitive statistical reasoning tasks.

The only two studies investigating signatures in a statistical reasoning task in human infants used looking-time patterns and produced mixed results. On the one hand, Téglás, Ibanez-Lillo, Costa and Bonatti (2015) found that, after watching a scene containing moving objects of two ensembles, infants looked longer at an unlikely than at a likely single-case outcome when the scene depicted a 3:1 ratio. However, they did not do so with a 12:4 ratio, suggesting that absolute set sizes, rather than ratios, influenced performance in this intuitive statistical reasoning task. On the other hand, Kayhan, Gredebäck and Lindskog (2017) measured infants' looking patterns at two multi-item-samples drawn from one population. In order to vary the magnitude of difference in likelihoods between samples, they manipulated both the ratios within samples and within the population. Here, results showed that infants' looking patterns varied as a function of the magnitude of difference in likelihood, suggesting that ROR does modulate infants' probability estimations. Hence, to date it remains unclear whether or not the analogue magnitude system is the primary cognitive mechanism enabling intuitive statistics in humans.

Despite this controversy, there is some indirect evidence supporting the idea of the analogue magnitude system as foundation of statistical reasoning: Both developmental and comparative research have shown that one important prerequisite capacity for statistical reasoning – tracking relative frequencies – is subject to the same signatures as tracking absolute frequencies. McCrink and Wynn (2007) presented human infants with a ratio discrimination task: After habituating them with multiple examples of a single ratio, infants were able to discriminate between new examples of this ratio and novel ratios. Infants' accuracy was highly dependent on the ratio between ratios, in accordance with Weber's law. Similar results were found for a nonhuman primate species (Drucker, Rossa, & Brannon, 2015): two rhesus macaques were trained to choose arrays that contained the greater ratio of positive to negative stimuli. Subjects' performance was modulated by the ratio between ratios: they responded more quickly and accurately the higher the ratio between ratios was, regardless of the absolute number of stimuli within the arrays. Results of these two studies (Drucker et al., 2015; McCrink and Wynn, 2007) suggest that the ability to discriminate ratios is a function of the ratio between the ratios to be discriminated, similarly as the ability to discriminate absolute quantities is a function of the ratio of the absolute set sizes. This raises the question whether the analogue magnitude system not only enables an individual to *track* relative frequencies, but also to *use* relative frequency information to draw statistical inferences from population to sample and vice versa.

One recent study did find some such evidence for an involvement of the analogue magnitude system in decision making under uncertainty in nonhuman great apes: Hanus and Call (2014) gave chimpanzees the choice between two trays on which food items were hidden under cups. The trays differed with respect to the ratio of food items to cups and thus in chances of finding food. Results showed that chimpanzees' performance varied as a function of the ratio between the two ratios, even in conditions where one tray constituted a 100% likelihood of finding food. This suggests that the ratio between ratios, more than the magnitude of difference within the single ratios, is decisive for apes' ability to discriminate probabilities. However, chimpanzees' success in this study could be explained with an absolute quantity heuristic, not regarding the amount of food items available, but regarding the number of cups on each tray. In particular, the tray depicting the more favorable

food/cup ratio always held the smaller number of cups. The authors added a control condition in which they excluded simple associative learning explanations (subjects did not preferentially choose the tray with fewer cups when the food was visibly removed from all cups). Nevertheless, it cannot be excluded that, as soon as there was any food to be found, chimpanzees at least partially relied on a mental shortcut such as "fewer cups = higher likelihood of finding food".

In sum, nonhuman great apes share with humans the fundamental ability to draw statistical inferences from population to sample and (to a certain extent) vice versa. Yet, it still remains an open question what the cognitive foundations of intuitive statistics are and whether they are the same in humans and their closest living relatives. To date, two main reasons suggest that the analogue magnitude system is the most plausible candidate for a basic statistical inference mechanism. First, the prerequisite capacity for this kind of inferences, tracking ratios, shows the same signatures as absolute quantity discrimination, both in human and nonhuman primates. Second, decision making under uncertainty seems to be ratio dependent in chimpanzees. What is missing is a comparative study testing great apes and humans in an intuitive statistical reasoning task that systematically varies the ROR and, crucially, controls for the use of absolute quantity heuristics.

The rationale of the present study, therefore, was threefold. First, we wanted to elucidate the cognitive and evolutionary underpinnings of intuitive statistics in chimpanzees and humans by testing its signatures. If intuitive statistics are based on the same analogue magnitude system, we would expect the characteristic performance patterns and signatures in both species. Much like the discrimination of absolute set sizes varies as a function of the ratio of the set sizes, we expected that the accuracy of intuitive statistics would vary as a function of the ROR between sets. We were particularly interested in determining the ROR lower threshold where performance breaks down in each species. A comparison of these signature limits with those found in simple quantity discrimination tasks helped us determine whether the inclusion of the statistical operation adds error to the representation in comparison to basic quantity discrimination. As described earlier, quantities are represented in an analogue, approximate way. We assumed that discriminating ratios of quantities *and* forming probabilistic expectations on their basis adds considerable noise relative to absolute quantity discrimination, since it requires representing and operating on quantities over multiple accounts (see, e.g. Barth et al., 2006 for an example of how subtraction operations add error to the quantity representation in comparison to simple quantity discrimination tasks). Accordingly, we expected the ROR threshold to be higher (i.e. less sensitive) than the threshold for discriminating absolute quantities. To address these questions, we presented chimpanzees with a previously established paradigm (Rakoczy et al., 2014) that required them to reason from populations of food items with different ratios of preferred and non-preferred food items to randomly drawn samples. We systematically varied the ratio between the two ratios (of preferred to non-preferred food items; ROR) ranging from 1 (equal proportions in both populations) to 16 (high magnitude of difference between populations). We tested human adults in a computerized version of the same task.

Our second goal was to replicate and validate previous findings on intuitive statistics in great apes. So far, only one experimentally highly experienced population of chimpanzees (and other apes) has been tested for their statistical abilities (Rakoczy et al. 2014). Therefore, it remains an open question whether findings of this particular population are generalizable to chimpanzees as a species. To investigate whether intuitive statistics is in fact a common, natural capacity in chimpanzees, we used the same task setup to test completely naïve individuals who were raised and housed in a different environment. This allowed us to directly compare between the previously tested, captive born and zoo housed chimpanzees, and the wild born, sanctuary housed chimpanzees tested in the present study.

Lastly, our study rules out alternative explanations based on absolute quantity heuristics. Most importantly, we included a crucial

_____
(footnote continued)
1999; Underwood, 1997; Zar, 1999). The only legitimate statement this study can draw regarding ROR effects is, therefore, that the effect was only visible in certain age groups depending on whether or not absolute and relative frequencies were confounded.

experiment to control for the possibility that apes used a strategy based on avoiding the population with the largest number of non-preferred items. The to date only study on nonhuman primates addressing such an avoidance heuristic has produced negative results: Capuchin monkeys failed to choose the sample from the proportion wise favorable population when they could not rely on the absolute amount of non-preferred items (Tecwyn, Denison, Messer, & Buchsbaum, 2016). It is, therefore, crucial to explore the possibility that apes' success in the present and previous studies was due to this simple quantity heuristic. Moreover, our study design ensured that apes could not succeed by choosing based on the absolute quantity of preferred food items. While previous studies (Eckert et al., 2017; Rakoczy et al., 2014) addressed this issue in separate control conditions, we designed our study in a way that apes were prevented from using such a strategy in *all* test conditions. Lastly, we also controlled for heuristics dealing with the total amount of food in each population. Hence, this is the first study on intuitive statistics in great apes comprehensively controlling for absolute number heuristics both regarding preferred and non-preferred items as well as their absolute total amount.

2. Methods

2.1. Chimpanzees

2.1.1. Subjects

We tested 24 chimpanzees (12 females) aged between 9 and 32 years at Ngamba Island Chimpanzee Sanctuary, Uganda. All individuals were born in the wild and orphaned at young age before they were transferred to the sanctuary. Research strictly adhered to the legal requirements of Uganda and was approved and reviewed by the Ugandan Wildlife Authorities and the Ugandan National Council for Science and Technology. The study was approved by the ethics committees of the Max Planck Institute for Evolutionary Anthropology and the Chimpanzee Sanctuary & Wildlife Conservation Trust.

2.1.2. Design and procedure

Subjects were tested in eight ROR conditions and one condition controlling for the usage of a non-preferred food avoidance heuristic (hereafter: carrot avoidance control; see Fig. 2). All conditions consisted of 12 trials, presented in a single session. The sequence of conditions was randomized for each subject, with the exception of the carrot avoidance control, which was the fifth condition for all subjects (see SI for more details).

We presented subjects with two transparent buckets filled with mixed populations of peanuts and carrot pieces of roughly equal size and shape. In all but one condition (ROR 1), one of the buckets contained a population that was more favorable in terms of its proportion of peanuts to carrots compared to the other. The experimenter showed both buckets to the subject, directed her gaze towards the ceiling and drew one item (always of the majority type) out of each of the buckets in a way that the chimpanzee could not see what was drawn. The experimenter kept the items hidden in her fists and the subject was allowed to indicate a choice to receive the chosen sample. In half of the trials, the experimenter crossed her arms when moving the fists towards the mesh to ensure that subjects made a choice between samples and not just chose the side where the favorable population was still visible. Trials with and without crossing were alternated (see Fig. 1 for an illustration).

2.1.2.1. ROR conditions.
To find the signatures of intuitive statistics, we systematically varied the ratio between populations ratios (ROR) ranging from 1 (equal ratio in both populations) to 16 (high magnitude of difference between ratios). In all ROR conditions we disentangled absolute and relative frequencies of peanuts: the bucket containing the less favorable ratio of peanuts to carrots contained twice as many peanuts as the bucket containing the more favorable ratio (see
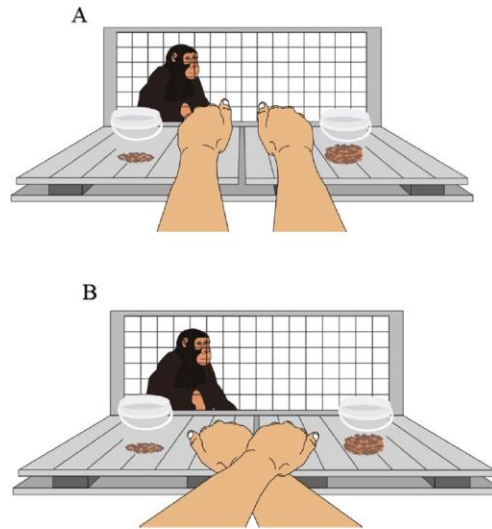
Fig. 1. Experimental setup. The experimenter simultaneously drew one sample from each of two populations in a way that kept the object hidden from the chimpanzee. The experimenter then gave the subject a choice between the two hidden samples (A). In half of the trials the experimenter crossed her arms before the subject was given a choice (B).

Fig. 2). This ensured that subjects truly reasoned about proportional information, rather than about absolute numbers of peanuts. RORs were calculated in the following way (following Drucker, Rossa, & Brannon, 2016; for more details see SI):

$$\frac{\text{Ratio of peanuts to carrots in the more favorable population}}{\text{Ratio of peanuts to carrots in the less favorable population}}$$

One consequence of this was that the favorable population always contained a smaller total amount of food items across ROR conditions (see Fig. 2). Hence, it is theoretically possible that subjects learned over the course of sessions to always pick the sample from the bucket with fewer items. Although this seems unlikely considering that chimpanzees have proven to be poor in learning arbitrary associations (e.g., Call, 2004), we included a ROR 1 condition in which both populations contained the same proportion of peanuts to carrots (i.e. both of them were equally likely to lead to a peanut as randomly drawn sample), while one contained double the amount of food items in absolute terms. If they still preferred the sample of one of the two populations, this would suggest that chimpanzees used other information than intended to solve the task. The reward pattern in this condition was adjusted accordingly: In randomized order, the experimenter drew a peanut from both buckets in half of the trials; in the other half she drew a carrot.

2.1.2.2. Carrot avoidance control.
To test whether subjects solved the task using a non-preferred food avoidance strategy focusing on the absolute number of carrots, here the favorable population contained four times more carrot pieces than the unfavorable population, while the ROR was 16 (see Fig. 2). If chimpanzees used a carrot avoidance strategy, we expected them to perform worse than in the ROR 16 condition.

2.1.3. Coding and data analysis

The apes' choice was coded live by the experimenter. A second blind observer coded 25% of the trials from video. Both raters were in

Fig. 2. List of conditions (eight ROR conditions and carrot avoidance control) and the respective proportions within the populations as well as the likelihood of drawing a peanut as a sample. Numerals in front of the colon depict numbers of preferred items, numerals after the colon depict numbers of non-preferred items. The first line always displays the favorable population except in ROR 1 in which both proportions were identical.

excellent agreement (K = 0.95, N = 576). To investigate whether there was an effect of the ROR on chimpanzees' performance across ROR conditions, we ran a Generalized Linear Mixed Model (GLMM; Baayen, 2008). "Correct choice" (choice of sample from population with higher peanut proportion) was the dependent variable. Note that data for the ROR 1 condition was not included in the model, since there was no "correct choice" in this condition (both populations were equally likely to lead to a preferred food item as a reward). As fixed effects we included logROR (since we expected the effect to be logarithmic), session and trial number (to test for potential learning effects) as well as all second and third order interactions among logROR, session number and trial number. To control for the effect of age and age^2 (in case of a nonlinear age-effect) they were included as further fixed effects. Subject ID and session ID were included as random effects. To keep type I error rate at the nominal level of 5% (Schielzeth & Forstmeier 2009; Barr 2013) we included all possible random slopes components (logROR, session number, trial number within subject ID and trial number within session ID) and also the respective correlations between random slopes and intercepts (see SI for more details on the statistical analysis). In order to determine the ROR lower threshold where performance breaks down, we inspected the confidence interval limits of the model: The model predicts performance to be above chance level in conditions for which the confidence interval limits lie above 0.5.

To investigate whether chimpanzees' performance in the carrot avoidance control was different from the ROR 16 condition, we ran a second GLMM. Again, "correct choice" was the response variable. As fixed effects we included condition (carrot avoidance control vs. ROR 16), session number and trial number as well as the two-way interactions between condition and trial number and between session number and trial number. Again, we included age and age^2 as further fixed effects, subject ID and session ID as random effects, and all random slopes components (condition, session number and trial number within subject ID and trial within session ID) as well as the respective correlations between random slopes and intercepts (see SI for more details).

2.2. Humans

2.2.1. Subjects

We tested 144 adult humans (80 women) aged between 18 and 34 years at the University of Göttingen, Germany. Participants were tested in a computerized version of the same task as the chimpanzees.

On a test computer, they were invited to imagine collecting as many red balls as possible from pairs of transparent urns filled with red and blue balls. They were asked to envision drawing from one urn of each pair with eyes closed and to indicate their choice by pressing one of two keys as quickly as possible.

2.2.2. Design and procedure

Similar to the apes, humans were tested in eight ROR conditions ranging from 1 to 16, with proportions resembling those for chimpanzees (see Fig. 2). To prevent participants from learning to always pick the urn with the smaller absolute quantity (since humans were tested in a single session the likelihood for learning such a rule was high), we also tested seven additional conditions (RORs between 1.5 and 16) in which absolute and relative frequencies were confounded. Since we expected getting a ceiling effect for confounded conditions (and their interpretation would not have been meaningful in terms of statistical reasoning), we did not plan to focus on these conditions in the analysis (but see SI for results). To prevent participants from counting the balls, images were only displayed for a maximum of 4 s (see SI for an example stimulus). Participants saw six trials per condition and trials of all conditions were randomized for each subject in one single test session.

2.2.3. Coding and data analysis

Participants' choice and response time were recorded automatically by EPrime (mean response times are depicted in SI Fig. 2). We used the same analysis as for chimpanzees, with the following exceptions: We ran two separate models, one for disentangled RORs and one for confounded RORs. Since humans were tested in a single session, we did not include session number as fixed effect, nor did we include session ID as random effect. Further, we only included age, but not age^2 as fixed effect, since we tested a small age range and did therefore not expect a nonlinear effect. Due to a significant effect of trial number, we ran the model for disentangled RORs again with only trial 1 performance considered (see SI for further details and results of the trial 1 model).

3. Results

3.1. Chimpanzees

3.1.1. ROR conditions

Chimpanzees performance was significantly influenced by the

Table 1

Influence of logROR, session, trial, age and age^2 on chimpanzees' proportion of correct choices.

Term	Estimate	SE	X^2	Df	P
Intercept	0.60	0.11	(1)	(1)	(1)
logROR[2]	0.21	0.05	15.44	1	p < 0.001
Session[2]	0.12	0.07	2.65	1	0.104
Trial[2]	−0.02	0.05	0.19	1	0.667
Age[2]	0.03	0.09	0.10	1	0.750
(Age)^2[2]	−0.08	0.06	2.06	1	0.151

Note: [1] not shown because lacking a meaningful interpretation; [2] these predictors were z-transformed.

logROR (GLMM, estimate ± SE = 0.21 ± 0.05, X^2 = 15.44, df = 1, P < 0.001; see Table 1 and SI for details), i.e. performance increased as a function of the ROR (see Fig. 3A) from a mean of 56.9% correct trials in ROR 1.5 to 69.8% in ROR 16 (see Table 2). The mean number of correct trials in ROR 1 was 51.4%. The model predicted the limit of chimpanzees' abilities to be between ROR 2 and 4 (see Fig. 3A). There was no effect of session or trial number, suggesting that chimpanzees' performance did not change with increasing experience within a session or over the course of sessions (see Table 1 and SI for more detailed information).

3.1.2. Carrot avoidance control

We found that subjects performed significantly better in the carrot avoidance control compared to the ROR 16 condition (GLMM, estimate ± SE = 0.78 ± 0.25, X^2 = 9.44, df = 1, P = 0.0016; see Fig. 3A and SI for details), which suggests that they did not use a strategy based on avoiding the population with more non-preferred food items.

3.2. Humans

Humans' performance was significantly influenced by the logROR (GLMM, estimate ± SE = 1.36 ± 0.08, X^2 = 163.53, df = 1, P < 0.001; see Table 3 and SI for details): Performance increased logarithmically as a function of the ROR (see Fig. 3B) from a mean of

43.4% correct trials in ROR 1.5 to 80.0% in ROR 16 (see Table 2 and Fig. 3B). The model predicted the same limit as for chimpanzees. We found an effect of trial number (X^2 = 4.58, df = 1, P = 0.029) indicating that participants' performance slightly increased with increasing experience. However, significance of the overall results did not change when considering trial 1 performance only (see SI for more information).

4. Discussion

We found that chimpanzee and human performance in a task requiring inferences from population to sample varied systematically as a function of the ROR, i.e., the magnitude of difference between the ratios of two populations. In accordance with Weber's law, performance increased logarithmically as a function of the ROR. This is the first piece of evidence to suggest that the analogue magnitude system is involved in intuitive statistical reasoning in both species. Intriguingly, and although methods for both species were somewhat different (e.g. live demonstration for chimpanzees vs. computer setup for humans) chimpanzees and humans displayed the same approximate limit: While the model predicted above chance level performance for ratios that differed by a factor of 4, it predicted failure for those that differed by a factor ≤ 2. Given that this is the first study addressing this topic and considering the small sample size, these results should be treated with caution. Nonetheless, the rather high threshold may suggest that, in comparison to simple quantity discrimination tasks, the statistical operation adds some error to the representation. Human adults, for instance, are able to discriminate absolute set sizes that differ by a factor of 1.15 (e.g. Barth, Kanwisher, and Spelke, 2003; Pica, Lemer, Izard, & Dehaene, 2004) compared to 6-month old infants who can discriminate ratios > 1.5 (Lipton & Spelke, 2003; Xu & Spelke, 2000). Not many studies documented the threshold for absolute quantity discrimination in nonhuman primates. Reported limits range from values as low as 0.9 for great apes (Hanus and Call, 2007) to 1.25 for rhesus macaques (Brannon & Terrace, 2000). The finding of a higher threshold for ratio discrimination within the realms of statistical inferences relative to basic quantity discrimination is consistent with the idea that additive error is to be expected when an organism represents and operates over multiple amounts (see, e.g. Barth et al., 2006; McCrink & Wynn, 2007
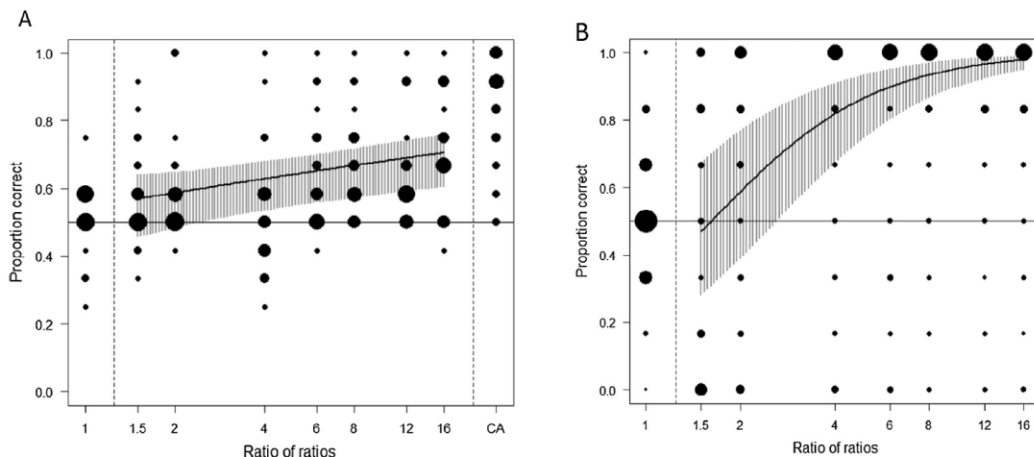

Fig. 3. Chimpanzee (A) and human (B) performance across conditions. Shown is the proportion of trials in which subjects chose the sample from the favourable population for all (disentangled) ROR conditions (in ROR 1 proportion of trials in which subjects chose sample from smaller population) and, for chimpanzees, the carrot avoidance control (CA). The dot size indicates the number of subjects performing at the same level (for readability purposes we used a different scale for the number of subjects in humans and chimpanzees). The horizontal line depicts chance level. The solid diagonal and vertical lines indicate the fitted model and its confidence limits. The model predicts performance to be above chance level in conditions for which the confidence interval limits lie above 0.5.

Table 2

Mean percentage of correct choices for each condition. *In ROR1 was no correct answer; here the percentage depicts the mean proportion of trials in which subjects chose the sample drawn from the population with the smaller absolute number of items.

Species		ROR1	ROR1.5	ROR2	ROR4	ROR6	ROR8	ROR12	ROR16	Carrot Avoidance
Chimpanzees	Mean proportion correct choices	51.4%*	56.9%	59.0%	53.1%	63.4%	66.2%	64.2%	69.8%	81.9%
Humans	Mean proportion correct choices	51.9%*	43.4%	59.8%	69.3%	72.1%	78.6%	79.5%	80.0%	/

Table 3

Influence of logROR, trial number and age and on humans' proportion of correct choices.

Term	Estimate	SE	X^2	df	P
intercept	1.96	0.27	(1)	(1)	(1)
logROR[2]	1.36	0.08	163.53	1	p < 0.001
Trial[2]	0.16	0.07	4.59	1	0.029
Age[2]	−0.32	0.27	1.36	1	0.239

Note: [1] not shown because lacking a meaningful interpretation; [2] these predictors were z-transformed.

for discussion of this hypothesis). Future studies with other populations of chimpanzees and humans, as well as different absolute quantities will have to examine whether we are truly dealing with a universal signature limit of statistical reasoning abilities. While chimpanzee performance was still far from ceiling even at the highest tested RORs, human performance rapidly increased and reached a plateau at ROR 8. This resembles findings on absolute quantity discrimination, where humans showed higher accuracy compared to other primates (e.g. Cantlon & Brannon, 2007), therefore suggesting that the same cognitive mechanism is utilized in both types of tasks. Future studies should test chimpanzees with a wider range of RORs to investigate whether (and when) they, like humans, also reach a maximum performance plateau.

One question that the present findings raise is whether intuitive statistics is based on an analogue magnitude system that is potentially not restricted to numerosity, but extends to all kinds of magnitudes, or alternatively on an approximate number system (ANS) in the more narrow sense, restricted to numerosity alone. This is related to the broader debate of whether numerical cognition is a quintessential cognitive domain with a specialized cognitive mechanism or whether the ANS is part of a broad domain in which all quantitative dimensions share computational mechanisms (see, e.g. Cantlon, Platt, & Brannon, 2009; Lourenco, 2015 for reviews). One way to inform this debate is to develop an intuitive statistical inference task disentangling discrete from continuous quantity information.

Another important question is whether the signatures of intuitive statistics and their limits extend to differences across life-stages known to apply in the case of the ANS with regard to absolute set sizes. More specifically, do we find better accuracy in older compared to younger individuals (analogous to findings by Halberda & Feigenson, 2008)? And are early inter-individual differences in accuracy predictive of later explicit statistical reasoning, as it is the case regarding the ANS with respect to absolute set sizes and later mathematical achievement (for a review see, e.g. Feigenson, Libertus, & Halberda, 2013)?

Our study not only gives insights into the cognitive foundation of intuitive statistics, the results also replicate those from our previous study (Rakoczy et al. 2014) in which we tested chimpanzees (and other nonhuman great apes) with substantial experience with cognitive testing. In that study, we had included a critical condition in which absolute and relative frequencies of preferred items were disentangled while keeping the ROR at a value of 16 – similar as in the present ROR 16 condition. While chimpanzees in Rakoczy et al. (2014) chose the sample from the favourable population in 66% of trials, chimpanzees in our present study performed correct in 69.8% of trials, even slightly exceeding the previously tested individuals' performance. Therefore, our ROR 16 condition replicated previous findings with a new

population of chimpanzees that was completely naïve to this kind of task and was raised and housed in a different environment (wild born and sanctuary housed vs. captive born and zoo housed). This suggests that intuitive statistical reasoning is a natural capacity in chimpanzees and not restricted to a single population with extensive experimental experience. Recent studies with two different monkey species (capuchin monkeys: Tecwyn et al., 2016; long-tailed-macaques: Plací, Eckert, Rakoczy, and Fischer, unpublished) using the same test paradigm failed to find unambiguous evidence for the presence of intuitive statistical abilities in levels comparable to apes.[3] Although much more research is needed, this might indicate that statistical reasoning has emerged late in primate evolution, perhaps only in the ape lineage. It would be of great interest to investigate the ecological pressures that could have led to the evolution from quantity discrimination abilities to probabilistic reasoning capacities.

Importantly, and in contrast to previous studies, our test design comprehensively controlled for the usage of simple heuristics based on absolute numbers of food items. In all ROR conditions we disentangled absolute and relative frequencies of peanuts, excluding the possibility that chimpanzees succeeded by simply picking the sample from the population with absolutely more preferred items. Crucially, the carrot avoidance control, which had not been tested in previous studies with apes, revealed that chimpanzees did not simply avoid the population containing more non-preferred food items (by contrast, capuchin monkeys did not perform significantly above chance level in a comparable condition in Tecwyn et al., 2016). Moreover, our ROR 1 condition provided evidence that chimpanzees truly used proportional information to solve the task, rather than, e.g. a "choose the sample from the bucket with less food"-strategy. Lastly, we did not detect any effect of session or trial number on chimpanzees' performance, making it unlikely that subjects used strategies learned over trials.

In sum, this study revealed that the signatures of intuitive statistics in chimpanzees and humans closely resemble those found in quantity discrimination tasks, thus strongly suggesting that these two abilities share the same basic and evolutionary ancient cognitive foundation, the analogue magnitude system. Moreover, we replicated previous findings on statistical reasoning in great apes with a new population of chimpanzees with a different housing and rearing background, suggesting that intuitive statistics is in fact a common capacity in chimpanzees. Lastly, this is the first study on intuitive statistics in great apes controlling for absolute number heuristics both regarding preferred and non-preferred items as well as absolute total amount, providing further evidence for true intuitive statistical reasoning in chimpanzees.

---

[3] On the group level, long-tailed macaques were only successful in conditions in which they could rely on a quantity heuristic dealing with the absolute number of preferred food items (e.g. 64:16 vs. 16:64). They failed in conditions, in which absolute and relative frequencies were disentangled (e.g. 48:12 vs. 12:192; Placi et al., unpublished). Capuchin monkeys, by contrast, succeeded even in conditions disentangling absolute and relative frequencies of preferred food items. Their performance in a non-preferred food avoidance control, however, only reached marginal significance after individuals, who exhibited a side-bias in this condition were removed from the analysis (Tecwyn et al., 2016).

Conflict of interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cognition.2018.07.004.

References

Ain, S. A., Giret, N., Grand, M., Kreutzer, M., & Bovet, D. (2009). The discrimination of discrete and continuous amounts in African grey parrots (Psittacus erithacus). *Animal Cognition, 12*(1), 145–154. https://doi.org/10.1007/s10071-008-0178-8.
Barnard, A. M., Hughes, K. D., Gerhardt, R. R., Divincenti, L., Jr., Bovee, J. M., & Cantlon, J. F. (2013). Inherently analog quantity representations in olive baboons (Papio anubis). *Frontiers in Psychology, 4*, 253. https://doi.org/10.3389/fpsyg.2013.00253.
Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics.* Cambridge: Cambridge University Press.
Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4*, 328. https://doi.org/10.3389/fpsyg.2013.00328.
Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86*(3), 201–221.
Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition, 98*, 199–222.
Beran, M. J. (2004). Chimpanzees (Pan troglodytes) respond to nonvisible sets after one-by-one addition and removal of items. *Journal of Comparative Psychology, 118*(1), 25–36. https://doi.org/10.1037/0735-7036.118.1.25.
Beran, M. J. (2017). Quantitative cognition. In J. Call (Ed.). *APA handbook of comparative psychology. Vol 2: Perception, learning, and cognition* (pp. 535–577). Washington, DC: American Psychological Association.
Bortz, J. (1999). *Statistik für Sozialwissenschaftler.* Berlin, Heidelberg: Springer.
Brannon, E. M., & Terrace, H. S. (2000). Representation of the numerosities 1–9 by rhesus macaques (Macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes, 26*(1), 31.
Buckingham, J. N., Wong, B. B. M., & Rosenthal, G. G. (2007). Shoaling decisions in female swordtails: How do fish gauge group size? *Behaviour, 144*, 1333–1346. https://doi.org/10.1163/156853907782418196.
Call, J. (2000). Estimating and operating on discrete quantities in orangutans (Pongo pygmaeus). *Journal of Comparative Psychology, 114*(2), 136–147. https://doi.org/10.1037//0735-7036.114.2.136.
Call, J. (2004). Inferences about the location of food in the great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, and Pongo pygmaeus). *Journal of Comparative Psychology, 118*(2), 232–241. https://doi.org/10.1037/0735-7036.118.2.232.
Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science, 17*(5), 401–406. https://doi.org/10.1111/j.1467-9280.2006.01719.x.
Cantlon, J. F., & Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biology, 5*(12), e328. https://doi.org/10.1371/journal.pbio.0050328.
Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Science (Regul. Ed.), 13*, 83–91. https://doi.org/10.1016/j.tics.2008.11.007.
Carey, S. (2009). *The origins of concepts.* New York: Oxford University Press.
Davidson, D. (1995). The representativeness heuristic and the conjunction fallacy effect in children's decision making. *Merrill-Palmer Quarterly (1982-), 328–346.
Dehaene, S. (2011). *The number sense: How the mind creates mathematics.* New York, NY: Oxford University Press.
De Neys, W., & Vanderputte, K. (2011). When less is not always more: Stereotype

knowledge and reasoning development. *Developmental Psychology, 47*(2), 432.
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 naive 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). 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. (2010b). Integrating physical constraints in statistical inference by 11-month-old infants. *Cogn Sci, 34*(5), 885–908. https://doi.org/10.1111/j.1551-6709.2010.01111.x.
Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition, 130*(3), 335–347.
Drucker, C. B., Rossa, M. A., & Brannon, E. M. (2016). Comparison of discrete ratios by rhesus macaques (Macaca mulatta). *Animal Cognition, 19*(1), 75–89. https://doi.org/10.1007/s10071-015-0914-9.
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.* 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.* https://doi.org/10.1016/j.cub.2018.04.077.
Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives, 7*(2), 74–79.
Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the 'number sense': the approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465.
Hanus, D., & Call, J. (2007). Discrete quantity judgments in the great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus): The effect of presenting whole sets versus item-by-item. *Journal of Comparative Psychology, 121*(3), 241–249. https://doi.org/10.1037/0735-7036.121.3.241.
Hanus, D., & Call, J. (2014). When maths trumps logic: Probabilistic judgements in chimpanzees. *Biology Letters, 10*(12), 20140892. https://doi.org/10.1098/rsbl.2014.0892.
Jacobs, J. E., & Potenza, M. (1991). The use of judgement heuristics to make social and object decisions: A developmental perspective. *Child Development, 62*(1), 166–178.
Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430–454.
Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237.
Kayhan, E., Gredebäck, G., & Lindskog, M. (2017). Infants distiguish between two events based on their relative likelihood. *Child Development.* https://doi.org/10.1111/cdev.12970.
Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science, 14*(5), 396–401.
Lourenco, S. F. (2015). On the relation between numerical and non-numerical magnitudes: Evidence for a general magnitude system. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.). *Mathematical cognition and learning. Vol 1: Evolutionary origins and early development of number processing* (pp. 145–174). London, San Diego, Waltham and Oxford: Elsevier.
McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychological Science, 18*(8), 740–745. https://doi.org/10.1111/j.1467-9280.2007.01969.x.
Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience, 32*, 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550.
O'Grady, S., Griffiths, T. L., & Xu, F. (2016). *Do simple probability judgements rely on integer approximation? Proceedings of the 38th annual conference of the cognitive science society.*
Piaget, J., & Inhelder, B. (1975). The origin of the idea of chance in children. (Trans L. Leake et al.).
Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science, 306*, 499–503.
Rakoczy, H., Cluver, A., Saucke, L., Stoffregen, N., Grabener, A., Migura, J., et al. (2014). Apes are intuitive statisticians. *Cognition, 131*(1), 60–68. https://doi.org/10.1016/j.cognition.2013.12.011.
Rugani, R., Cavazzana, A., Vallortigara, G., & Regolin, L. (2013). One, two, three, four, or is there something more? Numerical discrimination in day-old domestic chicks. *Animal Cognition, 16*(4), 557–564. https://doi.org/10.1007/s10071-012-0593-8.
Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology, 20*(2), 416–420. https://doi.org/10.1093/beheco/arn145.
Tecwyn, E. C., Denison, S., Messer, E. J., & Buchsbaum, D. (2016). Intuitive probabilistic inference in capuchin monkeys. *Animal Cognition.* https://doi.org/10.1007/s10071-016-1043-9.
Téglás, E., Ibanez-Lillo, A., Costa, A., & Bonatti, L. L. (2015). Numerical representations and intuitions of probabilities at 12 months. *Developmental Science, 18*, 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.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458.
Underwood, A. J. (1997). *Experiments in ecology.* Cambridge: Cambridge University Press.

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 of the United States of America, 105*(13), 5012–5015.

https://doi.org/10.1073/pnas.0704450105.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11.

Zar, J. H. (1999). *Biostatistical analysis* (4th ed.). New Jersey: Prentice Hall.

Appendix C: Eckert, Rakoczy, Call, Herrmann, & Hanus (2018)

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. doi: 10.1016/j.cub.2018.04.077

Current Biology

Chimpanzees Consider Humans' Psychological States when Drawing Statistical Inferences

Highlights

- Chimpanzees reasoned probabilistically from populations of food items to samples

- We varied whether experimenters were biased or neutral and had visual access or not

- Apes flexibly considered mental state information to judge randomness of events

- Further evidence for a shared statistical inference mechanism in apes and humans

Authors

Johanna Eckert, Hannes Rakoczy, Josep Call, Esther Herrmann, Daniel Hanus

Correspondence

eckert.johanna@gmail.com

In Brief

Eckert et al. show that chimpanzees, when drawing statistical inferences from population to sample, flexibly consider whether the experimenter is neutral or biased and whether she has visual access or draws blindly. This suggests that great apes possess a sophisticated intuitive statistical reasoning mechanism on a par with that of human infants.

CellPress

CellPress

Chimpanzees Consider Humans' Psychological States when Drawing Statistical Inferences

Johanna Eckert,[1,2,3,5,*] Hannes Rakoczy,[2,3] Josep Call,[1,4] Esther Herrmann,[1] and Daniel Hanus[1]
[1]Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
[2]Department of Developmental Psychology, University of Göttingen, Waldweg 26, 37073 Göttingen, Germany
[3]Leibniz Science Campus "Primate Cognition", German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany
[4]School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, UK
[5]Lead Contact
*Correspondence: eckert.johanna@gmail.com
https://doi.org/10.1016/j.cub.2018.04.077

SUMMARY

Great apes have been shown to be intuitive statisticians: they can use proportional information within a population to make intuitive probability judgments about randomly drawn samples [1, J.E., J.C., J.H., E.H., and H.R., unpublished data]. Humans, from early infancy onward, functionally integrate intuitive statistics with other cognitive domains to judge the randomness of an event [2–6]. To date, nothing is known about such cross-domain integration in any nonhuman animal, leaving uncertainty about the origins of human statistical abilities. We investigated whether chimpanzees take into account information about psychological states of experimenters (their biases and visual access) when drawing statistical inferences. We tested 21 sanctuary-living chimpanzees in a previously established paradigm that required subjects to infer which of two mixed populations of preferred and non-preferred food items was more likely to lead to a desired outcome for the subject. In a series of three experiments, we found that chimpanzees chose based on proportional information alone when they had no information about experimenters' preferences and (to a lesser extent) when experimenters had biases for certain food types but drew blindly. By contrast, when biased experimenters had visual access, subjects ignored statistical information and instead chose based on experimenters' biases. Lastly, chimpanzees intuitively used a violation of statistical likelihoods as indication for biased sampling. Our results suggest that chimpanzees have a random sampling assumption that can be overridden under the appropriate circumstances and that they are able to use mental state information to judge whether this is necessary. This provides further evidence for a shared statistical inference mechanism in apes and humans.

RESULTS

We used an established paradigm [1] in which chimpanzees faced two mixed populations of preferred and non-preferred food items and could choose from which of the two populations they wanted to receive a sample. In contrast to previous studies where drawing was always random, we here varied whether sampling was random or not (method adapted from [3]). To examine whether chimpanzees could integrate knowledge about others' choice biases and visual access into their statistical inferences, we first demonstrated to them that two experimenters, E1 and E2, had specific and opposing biases regarding two types of food in experiment 1: E1 preferred the type of food liked less by the apes themselves (carrot), whereas E2 showed the same preferences as the apes (peanut). These choice biases were established as follows: E1 repeatedly drew only carrots from a population with mostly peanuts (200 peanuts and 20 carrot pieces), and E2 showed the reverse patterns, repeatedly drawing only peanuts from a population with mostly carrots (20 peanuts and 200 carrot pieces). During the subsequent two test conditions, subjects witnessed the two experimenters sampling from their respective populations and were allowed to pick one of the samples. As the samples were hidden inside E1's and E2's fists, they had to infer from which population or experimenter they would most likely receive a favorable food item as a sample. The crucial manipulation between conditions was whether the experimenters looked into the bucket during sampling (visual access condition, see Figures 1A and 1B) or drew blindly (no visual access condition, see Figures 1C and 1D). The order of these two test conditions was counterbalanced across subjects.

To examine chimpanzees' baseline performance in this task without any prior information about experimenters' choice biases, we tested them in experiment 2 with new food types in the same proportions as before (200:20 versus 20:200). Similar to the no visual access test, both experimenters drew blindly from the populations.

We found that subjects' choice in experiment 1 was significantly influenced by conditions (GLMM; X^2 = 44.26, df = 1, p < 0.001). More specifically, in the visual access condition, when experimenters looked into the buckets, chimpanzees preferentially picked the sample drawn from the less favorable population (Mean$_{favorable\ population}$ = 33.8%), significantly
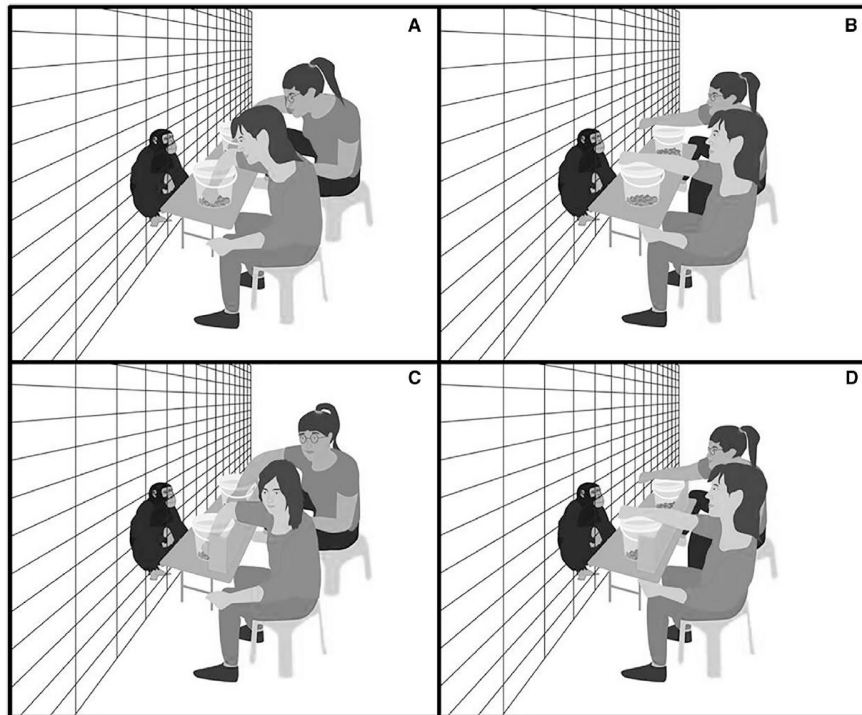
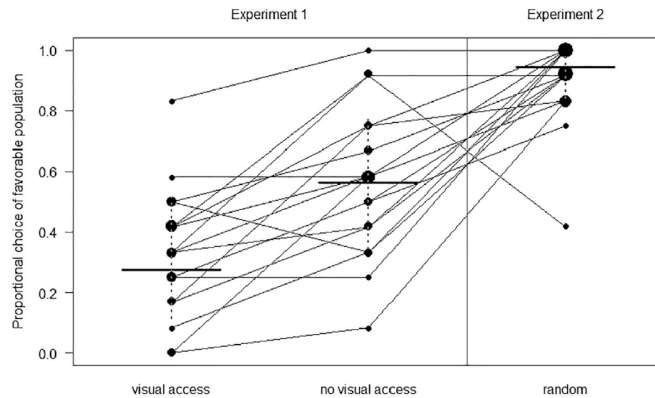**Figure 1. Illustration of the Procedures of the Two Test Conditions in Experiment 1**

(A–D) In the visual access condition, experimenters looked into the buckets while sampling (A) before offering the subject a choice between the two samples hidden in their fists (B). In the no visual access condition (C and D), a screen was placed between experimenters and buckets, blocking the experimenters' view into the populations. Moreover, in this condition, the experimenters' faces and bodies were oriented away from the table, further emphasizing a lack of visual access to the buckets during sampling.

different from what would be expected by chance (t = −3.58, df = 19, p = 0.002). Thus, subjects based their choice on the experimenters' choice biases rather than on the proportional composition of the population. In contrast, when the same experimenters sampled blindly in the no visual access condition, subjects' choice was different; here, they tended to choose the sample from the more favorable population more often, albeit not above what would be expected by chance (Mean$_{favorable\ population}$ = 57.1% of trials; t = 1.37, df = 19, p = 0.187), but a comparison of the two conditions revealed that subjects chose the proportionally favorable population significantly less often in the visual access condition than in the no visual access condition (estimate ± SE = −1.083 ± 0.204, df = 2, p < 0.001, confidence interval [CI] [−1.714, −0.496], see Figure 2). This pattern was not due to any order effects, since it held equally for both orders of presentation of the test conditions (X^2 = 0.007, df = 1, p = 0.931). Moreover, the effect was not driven by single individuals. Apart from one young female showing the opposite pattern and two subjects showing no difference between conditions, all remaining 17 individuals

chose the sample from the more attractive population numerically more often in the no visual access condition.

In experiment 2, when subjects did not have any prior information about potential choice biases and drawing was random, chimpanzees chose the sample from the more favorable population at the highest levels (Mean$_{favorable\ population}$ = 88.9% of trials), significantly above chance level (t = 11.78, df = 17, p < 0.001) and significantly more often than in the visual access condition (estimate ± SE = 3.261 ± 0.355, df = 2, p < 0.001, CI [2.416, 4.548]) and in the no visual access condition of experiment 1 (estimate ± SE = 2.177 ± 0.352, df = 2, p < 0.001, CI [1.234, 3.317]; see Figure 2).

We did not find any effect of trial number within the conditions for the two experiments, indicating that chimpanzees did not learn within a session which of the two populations or experimenters was the rewarded one (X^2 = 2.693, df = 2, p = 0.260). First trial performance confirmed the choice pattern: 45% of subjects chose the sample coming from the more attractive population in the first trial of the visual access condition compared to 60% in the no visual access condition and 78% in the random

Figure 2. Proportion of Trials in which Subjects Chose the Sample Drawn from the More Favorable Population in Experiments 1 and 2

Size of the dots represents number of subjects performing at the same level. Bold horizontal lines depict the mean probability predicted by the model, and gray dotted vertical lines depict bootstrapped 95% CIs.

See also Tables S2 and S3.

condition. The identity of the experimenter did not influence the chimpanzees' choice ($X^2 = 1.130$, df = 1, p = 0.264; see also Tables S2 and S3 for detailed results of experiments 1 and 2).

To control for potential associative learning explanations, we lastly tested chimpanzees in experiment 3, again using populations of new food types (100:10 versus 10:100). Before the test, subjects experienced that both experimenters would always draw preferred food items out of their population. However, while E1 sampled blindly from the more favorable population, E2 sampled from the less favorable one while looking into the bucket. In the subsequent test, both experimenters drew in the same manner as before but this time from identical populations containing equal proportions of preferred to non-preferred food items (55:55 versus 55:55). We found that chimpanzees preferred the sample drawn by the experimenter who had before sampled the statistically unlikely (preferred) food type significantly above chance level (Mean$_{favorable\ experimenter}$ = 64.8% of trials; t = 4.438, df = 17, p < 0.001; CI [0.577, 0.718]; see Figure 3). Again, we did not find an effect of trial number ($X^2 = 0.007$, df = 1, p = 0.933), indicating that subjects did not learn within the test session which experimenter was favorable (see also trial 1 performance: 66.7%). Moreover, we did not find an effect of experimenter's ID, neither when considering only experiment 3 ($X^2 = 0.803$, df = 1, p = 0.370) nor when considering whether it was the same or the opposite one compared to experiment 1 ($X^2 = 1.142$, df = 1, p = 0.286), indicating that subjects did not perform better when the positive experimenter was the same as in the first experiment (also see Table S4).

DISCUSSION

The current study shows that chimpanzees were able to flexibly adapt their choice as a function of statistical and psychological information in a paradigm that required them to reason probabilistically from population to sample. In the visual access condition of experiment 1, when biased experimenters drew samples while looking into the bucket, chimpanzees preferred the sample drawn by the experimenter with the preference for the favorable food type, mostly disregarding the proportional composition of the populations. This suggests that subjects expected the

drawing to be based on the experimenters' choice biases and, therefore, non-random in this condition. When the same biased experimenters drew samples from the same populations in the no visual access condition blindly, subjects switched and now showed a slight preference for the proportion-wise more favorable population despite the experimenters' biases. Hence, depending on whether or not the experimenters had visual access to the buckets while drawing, subjects based their choice either on the experimenters' choice biases or rather on the mere proportional composition of the population. In experiment 2, when chimpanzees did not have any information about potential biases of the experimenters and they drew blindly, subjects chose the sample drawn from the favorable population at higher levels than in both conditions of experiment 1. Results of these two experiments suggest that chimpanzees, without any prior information, assumed random sampling and expected the sample to reflect the population's distribution. If they, however, had reason to assume that the experimenters were biased, subjects' choice reflected these biases; the severity of this influence was dependent on whether the experimenters had visual access or not.

However, despite the fact that we did not find any indication for learning within test sessions, we cannot exclude that subjects might have learned during the demonstration of experiment 1 to simply associate one of the populations or experimenters positively or negatively and pick or avoid this one in the visual access condition where the setup was identical to the demonstration. The difference between conditions could congruently be explained by a change in setup in case of the no visual access condition (presence of a barrier) or the elapsed time in case of experiment 2. We believe this scenario is unlikely considering that chimpanzees and other nonhuman primates are known to have severe difficulties learning rules that clash with their natural predisposition to choose the larger of two (preferred) food amounts [7–9]. Furthermore, the shortness of the demonstration exposure makes any rule-learning explanation additionally implausible. Nevertheless, we sought to address this alternative explanation in experiment 3, in which chimpanzees were required to infer an experimenter's choice bias from statistical information (and according behavioral cues) without being differentially rewarded in the demonstration. In the test, subjects intuitively preferred the sample drawn by the experimenter who had previously drawn the statistically unlikely (and preferred) food type in the demonstration over the experimenter who drew blindly (and therefore randomly). This suggests that
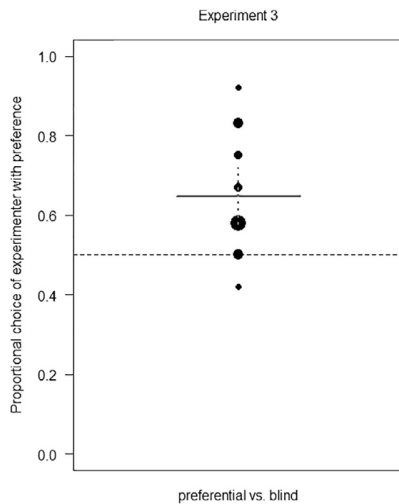
Figure 3. Proportion of Trials in which Subjects Chose the Sample Drawn by the Experimenter with a Bias for the Preferred Food Type, instead of the Sample Drawn by the Blind Experimenter, in Experiment 3

Size of the dots represents number of subjects performing at the same level. Bold horizontal line depicts the mean value for all subjects, and gray dotted vertical lines depict 95% CIs. Dashed horizontal line indicates chance level (i.e., indifference between both experimenters).
See also Table S4.

chimpanzees were able to use statistical information, in particular a violation of statistical likelihoods, to infer an experimenter's choice biases and draw conclusions about the sampling process. At the same time, it corroborates our hypothesis that subjects do not rely on associatively learned rules in this kind of task. It should be noted that, even though there is evidence that great apes have some understanding about human preferences or desires [10], we do not intend to make any strong claims about how chimpanzees interpreted the experimenter's choice bias in our study. It is possible for example that the subjects inferred that experimenter 1 seems to like (drawing or giving) peanuts. It is similarly possible that they simply noticed that experimenter 1, for whatever reason, draws peanuts when she has the possibility to do so. We cannot disentangle these two possibilities, and for the interpretation of our data, it is sufficient to assume the latter option.

While chimpanzees showed a remarkable flexibility and sophistication in this study, one may wonder why they did not perform better in the no visual access condition of experiment 1. Subjects in this condition chose the sample of the proportionally attractive population on average in 57% of trials as compared to 89% in experiment 2, although we used the exact same proportions in both experiments. The most likely explanation for this difference is that chimpanzees in experiment 2 did not have any information about potential biases of the experimenters, which left the randomness of the draw the only aspect to consider (results of this experiment also demonstrated that

subjects had not remembered any "good" or "bad" labels for the experimenter from the previous experiment). By contrast, in the no visual access condition of experiment 1, chimpanzees had to overcome what they had just experienced—namely, that E1 always extracted carrots from the peanut population, and E2 always extracted peanuts from the carrot population. This information was even repeated (in reminder trials) right before the no visual access condition. Thus, this condition required two extra steps compared to experiment 2: (i) chimpanzees had to recognize and understand the indicators of blind drawing, and (ii) they had to weigh the indicators of "biased sampling" and "blind sampling" against each other and choose accordingly. Therefore, our task design required a fair amount of cognitive flexibility, which might have been too demanding for some of the subjects.

Conclusions

Taken together, the results of our three experiments suggest that chimpanzees did consider information about the experimenters' choice biases and visual access when drawing statistical inferences. Subjects were not only able to recognize that sampling would be non-random when biased experimenters had visual access while drawing, but they also knew, to some extent, that when visual access was blocked, the choice bias information was rendered irrelevant and could therefore be dismissed. Moreover, chimpanzees were able to draw inferences about the experimenter and the sampling process from the given statistical information even without being differentially rewarded; when samples were unambiguously non-representative of a populations' distribution and the experimenter looked into the population while sampling, subjects seemed to infer that the sampling person must have a bias for drawing one of the food types and acted accordingly in the test condition. While previous studies have shown that chimpanzees can reason probabilistically from population to sample [1, J.E., J.C., J. Hermes, E.H., and H.R., unpublished data] and are sensitive to what others can and cannot see (both conspecifics [e.g., 11] and human experimenters [e.g., 12]), our study is the first to suggest that chimpanzees are able to flexibly combine these two sources of information to make rational decisions under uncertainty. Our results resemble findings on human infants; just as the chimpanzees in our study, 11-month-old infants were shown to be sensitive to whether a sample was or was not drawn randomly from a population on the basis of information about the drawing agent's psychological states (her preference and visual access) [3]. Similar to our apes, infants were also able use statistical information (in particular a violation of likelihoods) to draw conclusions about the sampling agent and the sampling process [13, 14].

Our study therefore gives further reason to assume that human statistical reasoning might be grounded in a cognitive mechanism that is utilized from early infancy onward and shared with our closest living relatives.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

SUPPLEMENTAL INFORMATION

Supplemental Information includes four tables and can be found with this article online at https://doi.org/10.1016/j.cub.2018.04.077.

AUTHOR CONTRIBUTIONS

Conceptualization, D.H., J.E.; Methodology, D.H., J.E.; Investigation, J.E.; Formal Analysis, J.E.; Writing—Original Draft, J.E.; Writing—Review & Editing, J.E., D.H., H.R., J.C., E.H.; Supervision, H.R., J.C.; Funding Acquisition, H.R., J.C., J.E.; Resources, E.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Rakoczy, H., Clüver, A., Saucke, L., Stoffregen, N., Gräbener, A., Migura, J., and Call, J. (2014). Apes are intuitive statisticians. Cognition *131*, 60–68.

2. Téglás, E., Girotto, V., Gonzalez, M., and Bonatti, L.L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. Proc. Natl. Acad. Sci. USA *104*, 19156–19159.

3. Xu, F., and Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. Cognition *112*, 97–104.

4. Denison, S., and Xu, F. (2010). Integrating physical constraints in statistical inference by 11-month-old infants. Cogn. Sci. *34*, 885–908.

5. Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J.B., and Bonatti, L.L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. Science *332*, 1054–1059.

6. Denison, S., Trikutam, P., and Xu, F. (2014). Probability versus representativeness in infancy: can infants use naïve physics to adjust population base rates in probabilistic inference? Dev. Psychol. *50*, 2009–2019.

7. Boysen, S.T., and Berntson, G.G. (1995). Responses to quantity: perceptual versus cognitive mechanisms in chimpanzees (Pan troglodytes). J. Exp. Psychol. Anim. Behav. Process. *21*, 82–86.

8. Schmitt, V., and Fischer, J. (2011). Representational format determines numerical competence in monkeys. Nat. Commun. *2*, 257.

9. Vlamings, P.H., Uher, J., and Call, J. (2006). How the great apes (Pan troglodytes, Pongo pygmaeus, Pan paniscus, and Gorilla gorilla) perform on the reversed contingency task: the effects of food quantity and food visibility. J. Exp. Psychol. Anim. Behav. Process. *32*, 60–70.

10. Buttelmann, D., Call, J., and Tomasello, M. (2009). Do great apes use emotional expressions to infer desires? Dev. Sci. *12*, 688–698.

11. Hare, B., Call, J., and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? Anim. Behav. *61*, 139–151.

12. Melis, A.P., Call, J., and Tomasello, M. (2006). Chimpanzees (Pan troglodytes) conceal visual and auditory information from others. J. Comp. Psychol. *120*, 154–162.

13. Wellman, H.M., Kushnir, T., Xu, F., and Brink, K.A. (2016). Infants Use Statistical Sampling to Understand the Psychological World. Infancy *21*, 668–676.

14. Gweon, H., Tenenbaum, J.B., and Schulz, L.E. (2010). Infants consider both the sample and the sampling process in inductive generalization. Proc. Natl. Acad. Sci. USA *107*, 9066–9071.

15. Baayen, R.H. (2008). Analyzing Linguistic Data: A Practical Introduction to Statistics (Cambridge: Cambridge University Press).

16. Schielzeth, H., and Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. Behav. Ecol. *20*, 416–420.

17. Barr, D.J. (2013). Random effects structure for testing interactions in linear mixed-effects models. Front. Psychol. *4*, 328.

18. Field, A. (2005). Discovering statistics using SPSS (London: Sage Publications).

19. Fox, J., and Weisberg, S. (2011). An R Companion to Applied Regression, Second Edition (Thousand Oaks, CA: Sage). https://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

20. Dobson, A.J. (2002). An Introduction to Generalized Linear Models, Second Edition (Chapman and Hall).

21. Forstmeier, W., and Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. Behav. Ecol. Sociobiol. (Print) *65*, 47–55.

22. R Core Team (2016). A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).

23. Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. *67*, 1–48.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Pan troglodytes	Ngamba Island Chimpanzee Sanctuary and Wildlife Conservation Trust, Uganda	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Johanna Eckert (eckert.johanna@gmail.com).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

In total, we tested 21 chimpanzees (9 females) with estimated ages between 7 and 33 years (Mean = 18.7). Twenty of those participated in experiment 1, 18 participated in experiment 2 and 3 (see Table S1 for more detailed information about the subjects). All individuals were wild born orphans that lived in a social group of 49 individuals at Ngamba Island Chimpanzee Sanctuary, Uganda. In accordance with the recommendations of the Weatherall report 'The use of nonhuman primates in research' chimpanzees were allowed to roam freely on the 40 ha island covered with tropical rainforest during the day and voluntarily spend the night in seven interconnected sleeping rooms (approx. 140 m2) with regular feedings and water *ad libitum*. Subjects participated in the study on voluntary basis and were never food or water deprived. Research strictly adhered to the legal requirements of Uganda and was reviewed and approved by the Ugandan Wildlife Authorities and the Ugandan National Council for Science and Technology. The study was ethically approved by committees of the Max Planck Institute for Evolutionary Anthropology and the Chimpanzee Sanctuary & Wildlife Conservation Trust. Animal husbandry and research comply with the 'PASA Primate Veterinary Healthcare Manual' and the 'Guidelines for the Treatment of Animals in Behavioral Research and Teaching' of the Association for the Study of Animal Behavior.

METHOD DETAILS

The study consisted of three experiments. experiment 1 was conducted in May 2017, experiment 2 and 3 were conducted consecutively in January 2018. In all experiments, subjects were tested individually in their sleeping rooms and two experimenters E1 and E2 were seated at a table (L/W/H: 73cm/40cm/48cm) in front of the subject close to the mesh. As stimuli we used mixed populations of preferred and non-preferred food items that were presented in two transparent buckets (⌀ 21.5 cm; height 19 cm). Food items of both types were of roughly equal size.

Experiment 1

Experiment 1 consisted of three phases: the demonstration phase, the visual access condition and the no visual access condition. In all phases, E1 presented a transparent bucket filled with 200 peanuts and 20 carrot pieces (P1), E2 presented a bucket filled with 20 peanuts and 200 carrot pieces (P2). We knew from previous studies (e.g., Eckert et al., under revision) that all tested individuals clearly preferred peanuts over carrots; hence, P1 was considered the more attractive population. The identity of E1 and E2 was counterbalanced across subjects. While all individuals started with the demonstration phase, the order of presentation of the two test phases was counterbalanced to avoid potential order effects. All three phases were tested on consecutive days. Twenty subjects participated in this experiment.

Demonstration phase

In the demonstration phase subjects experienced that both experimenters had a bias to sample items of the minority type in their bucket, i.e., E1 for carrot pieces (from population P1) and E2 for peanuts (from population P2). In other words, E2 had the same preference as the subject, while E1 had the opposing preference. A demonstration trial started with the right experimenter presenting her bucket by shaking it, tilting it, and turning it around to give the subject a good overview about the content. Subsequently, she looked into her bucket, searched for three seconds using one hand and then visibly drew one item (of the minority type) and handed it to the subject. In the next trial, the left experimenter did the same with her bucket. In one session, subjects received ten demonstration trials per experimenter, with both experimenters sampling in alternating order. The side on which the experimenters were seated was counterbalanced. Chimpanzees received a total of two demonstration sessions on two consecutive days.

Test phase

Each test condition (visual access condition and no visual access condition) was administered in a single session consisting of 12 trials. The order of presentation of the two conditions was counterbalanced. Before a test session, each subject received three reminder trials per experimenter using the same procedure as in the demonstration trials.

Visual access condition: Each trial started with the right experimenter presenting her population by shaking, tilting, and turning the bucket. Then the left experimenter did the same with her bucket. Subsequently, E1 and E2 simultaneously looked into their bucket, searched for three seconds using one hand and drew one item each without the subject seeing which item they had extracted (see Figure 1A). Just as in the demonstration, both experimenters always sampled an item of the minority type, i.e., E1 drew a less favorable item (carrot) out of the more favorable population (mostly peanuts), E2 drew a more favorable item (peanut) out of the less favorable population (mostly carrots). Both experimenters kept the sample hidden in their fist and presented the closed fist to the subject (see Figure 1B). The subject then indicated a choice between the two samples by pointing to one of the fists and immediately received the chosen sample as reward. Again, the side on which E1 sat was counterbalanced.

No visual access condition: The procedure was the same as for the visual access condition with the following modification: After having presented their buckets with the populations, the experimenters placed an opaque screen (L/W/H: 60cm/15cm/37cm) in between themselves and the buckets thereby blocking their view into the buckets. To further emphasize their lack of visual access, experimenters' body orientation and gaze was directed away from the buckets during the sampling process (see Figure 1C). Next, both experimenters drew quickly and randomly the first item they could grasp in the bucket. Subsequently, the subject was offered a choice between the two hidden samples (Figure 1D).

Experiment 2

In experiment 2, subjects were tested in a single test condition and did not experience any demonstration beforehand (and accordingly no reminder-trials). In order to avoid carry-over effects from experiment 1, we used new types of food, their preference hierarchy was established in preference tests before and after the experiment (see Table S1 for more detailed information). The proportions within the populations remained the same as before (200:20 versus 20:200). The procedure was similar to the no visual access condition of experiment 2 with the following two modifications: 1. Experimenters were not assigned to one of the buckets across trials; instead, we counterbalanced the number of trials in which each experimenter presented and sampled from each of the populations. Thereby we hoped to minimize chances that subjects would form good/bad associations with the experimenters (while we were at the same time able to detect such potential effects statistically post hoc). 2. We did not use the barrier to indicate blind drawing; instead, the experimenters just turned away from the buckets and directed their gaze toward the ceiling. In doing so we wanted to examine whether these cues are sufficient for the apes to assume random drawing, which was important for the subsequent experiment 3. Two subjects changed preferences over the course of the experiment (showing the opposite preference in the food preference test after the experiment compared to before). Accordingly, their data were excluded from the analysis. One further subject that had participated in experiment 1 did not show any preference for one of the food types and was therefore not tested. In total, we included 18 subjects in the analysis.

Experiment 3

Experiment 3 consisted of two phases, the demonstration phase and the test phase, which were administered on consecutive days (only one subject did not enter the sleeping room the day after the demonstration phase and therefore had a one-day-break between demonstration and test). We again used new types of food in order to avoid carry over effects from the previous experiments (see Table S1). One individual had to be excluded because of lack of motivation; two further subjects could not be tested because they either did not enter the sleeping rooms within our data collection period or because they did not show a clear preference for one of the food types. In total, we included 18 subjects in our data analysis.

Demonstration phase

In the demonstration phase subjects experienced that E1 would blindly draw preferred items from the more favorable population (P1: 100 preferred and 10 non-preferred items), while E2 would intentionally draw preferred food items from the less favorable population (P2: 10 preferred and 100 non-preferred items). In each trial, consecutively, E1 turned away, directed her gaze toward the ceiling and drew one item quickly from her population and handed it over to the subject; E2 looked into her bucket and searched for three seconds before she handed over a preferred food item to the subject. Both experimenters always drew the preferred food type (except for one trial each for two subjects in which E2 accidentally drew a non-preferred item. Note that we ran a second analysis without these two subjects which did not change the significance of the results). Per session, subjects saw ten demonstration trials per experimenter, with both experimenters sampling in alternating order. For half of the subjects the identity of E1 and E2 remained the same as in experiment 1, for the other half identities were swopped, which allowed us to test for carry over effects in our analysis. The side on which experimenters were seated was counterbalanced and the experimenter on the right always started sampling. Chimpanzees received a total of two demonstration sessions on two consecutive days.

Test phase

Before a test session, each subject received three reminder trials per experimenter using the same procedure as in the demonstration trials. In the test trials both experimenters had the exact same population with the same proportion of preferred to non-preferred food items (55:55). Hence, both populations depicted a 50% chance of leading to a preferred food item as randomly drawn sample. Each trial started with the right experimenter presenting her population by shaking, tilting, and turning the bucket. Then the left

experimenter did the same with her bucket. Subsequently, E2 looked into her bucket and searched for three seconds, while E1 turned away, directed her gaze toward the ceiling and moved her arm over the bucket. Then, both experimenters simultaneously drew one item without the subject seeing which item they had extracted. Both experimenters kept the sample hidden in their fist and presented the closed fist to the subject. The subject then indicated a choice between the two samples by pointing to one of the fists. Here, E2 always sampled preferred items, while E1 drew truly randomly. Again, the side on which E1 sat was counterbalanced. Chimpanzees received a total of 12 test trials presented in a single session.

QUANTIFICATION AND STATISTICAL ANALYSIS

The apes' choice was coded live. A second blind observer coded 20% of the trials from video for each experiment. Both raters were in excellent agreement (experiment 1: K = 0.96, N = 97; experiment 2: K = 1, N = 48; experiment 3: K = 0.94, N = 48). To analyze experiment 1 and 2, we ran a Generalized Linear Mixed Model (GLMM) [15] with subject's choice (between populations) as dependent variable. As fixed effects we included condition, order of conditions, experimenter ID and trial number (to check for potential learning effects) as well as the three-way-interaction between condition, order and trial number. To control for a potential (linear or non-linear) effect of subjects' age, we included age and age^2 as further fixed effects. Subject ID was included as random effect. To keep type I error rate at the nominal level of 5% [16, 17] we included all possible random slopes components (condition, trial number and experimenter ID within subject ID) and the respective correlations between random slopes and intercepts. Trial number, age and age^2 were z-transformed (to a mean of zero and a standard deviation of one). Variance Inflation Factors (VIF) [18] were derived for a standard linear model excluding the random effects and interactions, using the function vif of the R-package car [19] and did not indicate collinearity to be an issue. We assessed model stability by comparing the estimates derived by a model based on all data with those obtained from models with the levels of the random effects excluded one at a time. This revealed the model was stable. The significance of the full model as compared to the null model (comprising only age, age^2 and the random effect subject ID) was established using a likelihood ratio test (R function Anova with argument test set to ''Chisq'') [20, 21]. p values for the individual effects were based on likelihood ratio tests comparing the full with respective reduced models (R function drop1). The model was fitted in R [22] using the function lmer of the R-package lme4 [23]. To assess whether the average performance of subjects in the different conditions was different from what would be expected by chance, we used two-tailed one-sample t tests, which were also administered in R.

To analyze experiment 3 we ran a second GLMM with subject's choice (between experimenters) as dependent variable. As fixed effects we included trial number (to check for potential learning effects) and experimenter ID in experiment 1 and 3, as well as the interaction between experimenter ID in both experiments (to check whether, e.g., individuals who had the same person as ''positive experimenter'' in both experiments performed better than those who had the opposite person). To control for a potential (linear or non-linear) effect of subjects' age, we included age and age^2 as further fixed effects. Subject ID was included as random effect. Again, we included all possible random slopes components (trial number within subject ID) and the respective correlations between random slopes and intercepts. Trial number, age and age^2 were z-transformed (to a mean of zero and a standard deviation of one). Again, Variance Inflation Factors (VIF) did not indicate collinearity to be an issue and model the model was found to be stable. The significance of the full model as compared to the null model (comprising only age, age^2 and the random effect subject ID) was again tested using a likelihood ratio test. p values for the individual effects were based on likelihood ratio tests comparing the full with respective reduced models (R function drop1). Again, the model was fitted in R using the function lmer of the R-package lme4 and to assess whether the average performance of subjects was different from what would be expected by chance, we used a two-tailed one-sample t test.

DATA AND SOFTWARE AVAILABILITY

Analysis-specific code and data are available by request to the Lead Contact.

Curriculum Vitae

Personal data

Name	Johanna Eckert
Date of birth	November 22nd, 1987
Place of birth	Würzburg
Citizenship	German
Marital status	unmarried

Education

Since 10/2014	**PhD student** Behavior and Cognition Georg-August-University of Göttingen, Germany
10/2012-09/2014	**M.Sc. Biology with honors** Specialization in behavioral biology, animal ecology and tropical biology Julius-Maximilians University of Würzburg, Germany
10/2009-09/2012	**B.Sc. Biology** Julius-Maximilians University of Würzburg, Germany

Practical Experience

01/2018 – 02/2018 04/2017 – 06/2017 10/2016 – 11/2016	**Research stays at Ngamba Island Chimpanzee Sanctuary, Uganda**
10/2016 – 12/2017	**Organizer: "Animal Cognition Reading Group"** Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany
04/2016	**Lecturer: Seminar "Comparative Psychology/Primate Cognition"** Otto-von-Guericke-Universität Magdeburg, Germany
Since 10/2014	**Dissertation: "The evolutionary roots of intuitive statistics"** Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany and Department of Developmental Psychology, University of Göttingen, Germany Advisors: Hannes Rakoczy, Josep Call
01/2014 – 09/2014	**Master thesis: "Social comparison in long-tailed macaques"** Cognitive Ethology Laboratory, German Primate Center in Göttingen, Germany Advisors: Julia Fischer, Johannes Spaethe

08/2013 – 10/2013	**Internship: "Vocal communication in African Elephants"** Elephant Sanctuary in Plettenberg Bay, South Africa and Department of Cognitive Biology, University of Vienna, Austria Advisor: Angela Stöger-Horwath
02/2013 – 04/2013	**Internship: "Cognitive Research with long-tailed macaques"** Cognitive Ethology Laboratory, German Primate Center in Göttingen, Germany Advisor: Christian Schloegl, Vanessa Schmitt
03/2012 – 06/2012	**Bachelor thesis "Visual learning in stingless bees"** Tropical Field Station in La Gamba, Costa Rica and Department of Behavioral Physiology and Sociobiology , University of Würzburg, Germany Advisor: Johannes Spaethe

Publications

Placi, S., **Eckert, J**., Rakoczy, H., & Fischer, J. (submitted). Long-tailed macaques (*Macaca fascicularis*) can use simple heuristics but fail at drawing statistical inferences from populations to samples.

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. doi: 10.1016/j.cub.2018.04.077

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. doi: 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). doi: 10.1002/ajp.22693

Schmitt, V., Federspiel, I., **Eckert, J**., Keupp, S., Tschernek, L., Faraut, L., Schuster, R., Michels, C., Bugnyar, T., Mussweiler, T., & Fischer, J. (2015). Do monkeys compare themselves to others? *Animal Cognition.* doi: 10.1007/s10071-015-0943-4

Spaethe, J., Streinzer, M., **Eckert, J**., May, S., & Dyer, A. (2014). Behavioural evidence of colour vision in free flying stingless bees. *Journal of Comparative Physiology A, 200*(6), 485-496. doi: 10.1007/s00359-014-0886-2

Talks and Posters

Invited talk at Department of Psychology, University of Cambridge, UK, 2017
  "The evolutionary roots of intuitive statistics"

Talk at Behaviour Conference Estoril, Portugal, 2017
„Signature limits of intuitive statistics in chimpanzees"

Talk at International Primatological Conference (IPS), Chicago, USA, 2016
"Do great apes use statistical information to draw conclusions about populations of food items?"

Poster at 5th Rovereto Workshop on Cognition and Evolution (CogEvo), Rovereto, Italy, 2016
"Ape statisticians: The extent of intuitive statistics in great apes" (Abstract Award Winner)

Poster at Behaviour Conference Cairns, Australia, 2015
"Do great apes truly engage in intuitive statistics?"

Talk at Second European Student conference on Behaviour and Cognition, Zurich, Switzerland, 2015
"Do great apes truly engage in intuitive statistics?"

Poster at CRC Symposium Evolution of Social Behaviour, Göttingen, Germany, 2014
"Social Comparison Processes in Old World Monkeys"

Scholarships and Memberships

10/2016 – 11/2016	Outgoing grant of Leibniz-ScienceCampus Primate Cognition to carry out
03/2016 – 06/2017	research in great ape sanctuary
03/2012 – 06/2012	DAAD promos scholarship (German Academic Exchange Service)
2018 to date	Membership in Orang-Utans in Not e.V.
2015 to date	Membership in Leibniz-ScienceCampus Primate Cognition,Göttingen, Germany
2014 to date	Membership in Chances for Nature e.V., Team Peru, Göttingen, Germany