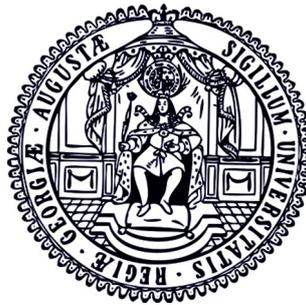


Statistical Reasoning in Nonhuman Primates and Human Children

Dissertation for the award of the degree
“Doctor of Philosophy (Ph.D.)”
Division of Mathematics and Natural Sciences
at the Georg-August-Universität Göttingen

in the doctoral program of Behaviour and Cognition (BeCog)
at the Georg-August University School of Science (GAUSS)



submitted by

Sarah Placi

from Neuchâtel (Switzerland)

Göttingen, 2019

Supervision

Prof. Dr. Julia Fischer (First Referee)

Cognitive Ethology Lab, German Primate Center, Göttingen

Prof. Dr. Hannes Rakoczy (Second Referee)

Developmental Psychology, Institute of Psychology, University of Göttingen

Prof. Dr. Michael R. Waldmann

Cognitive Science and Decision Making, Institute of Psychology, University of Göttingen

External thesis advisors

Dr. Tanya Behne

Developmental Psychology, Institute of Psychology, University of Göttingen

Prof. Dr. Annekathrin Schacht

Affective neuroscience and psychophysiology, Institute of Psychology, University of Göttingen

Dr. Dietmar Zinner

Cognitive Ethology Lab, German Primate Center, Göttingen

Day of the oral defence: 25.03.2019

Herewith I declare that I have written this thesis independently. Assistance of third parties was only accepted if scientifically justifiable and acceptable in regards to the examination regulation. All sources have been quoted.

Göttingen,

Sarah Placi

Published and submitted manuscripts

The present thesis is a publication-based dissertation. It is based on the following original research articles that have been published in, or have been submitted to international peer-reviewed journals.

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

Placì, S., Padberg, M., Rakoczy, H., & Fischer, J. (2019). Long-tailed macaques extract statistical information from repeated types of events to make rational decisions under uncertainty. *Scientific Reports*, 9(12107).

Placì, S., Fischer, J. & Rakoczy, H. (*submitted*). Do infants and preschoolers quantify probabilities based on proportions?

In the present dissertation I describe the theoretical environment that led to the conception of the studies presented in these articles and discuss how the main findings can be incorporated into the existing literature on the topic. I was the first author of all manuscripts. In particular, I was responsible for designing and conducting the experiments, for analyzing and interpreting the data, and for writing up, submitting, and publishing the manuscripts.

To my parents, Monika and Aldo, who kept the impossible at bay

To dreams and passion

Acknowledgements

I thank my two supervisors **Julia Fischer** and **Hannes Rakoczy** for having entrusted me with this special project. The investigation of statistical reasoning in different species opened a whole new world for me, filled with life-changing discoveries (and dark moments of doubts). I also thank them for letting me work independently and for the valuable information they allowed me to retrieve from their own wealth of knowledge. I learnt about much more than just my own topic by sharing parts of their scientific environment. I learnt about the importance of critical thinking and of skepticism in science, not so as to hinder progress but so as to foster careful and steady steps in which progress can safely tread. I learnt about the importance of opening one's mind to other topics more or less connected to one's field of research, to gain higher levels of abstraction and perceive meanings that would otherwise have been lost. I learnt much more from them, surely more than I am aware of, and for this I thank them very much.

I would also like to thank **Michael Waldmann** for having accepted to be part of my thesis committee and for having shared his wisdom with me during our meetings.

I thank **Tanya Behne**, **Annekathrin Schacht**, and **Dietmar Zinner** for having accepted to be members of the examination board.

I thank the **RTG2070** for its financial support, as well as all the helpful workshops it provided and the opportunity it gave me to visit the University of Berkeley in my last year of PhD. More particularly, I thank **Rebecca Jürgens** for her help regarding administrative questions.

I thank **Carolin Kade**, **Lukas Schad**, **Marie Padberg**, **Lydia Schidelko**, and **Isabelle Ganter** for their help in the collection of my data.

I thank **Lauren Cassidy**, **Johanna Eckert**, and **Simon Stephan** for having taken the time to go through my dissertation.

I thank all the members of my two labs, the Cognitive Ethology lab and the Department of Developmental Psychology, as well as all the members of the Research Training Group 2070, of the doctoral program Behaviour and Cognition, and of the German Primate Center, for having enriched my days with their friendliness and support. More particularly, I would like to thank **Federica Dal Pesco**, **Lauriane**

Faraut, and **Carolin Kade** for all the beautiful, funny, adventurous and intense moments we shared together. I also want to thank **Johanna Eckert**, **Johanna Prüfer**, **Lauren Cassidy**, **Alan Rincon**, **Adi Lausen**, **Clémence Poirotte**, **Louise Peckre**, **Sonia Tuitou**, **Matthis Drolet**, and **Delphine De Moor** for having filled these past three years and a half with laughter and friendship and for having contributed in making my PhD one of the biggest adventures of my life.

I would also like to thank other friends met at a different time and in other places of the world, but whose friendship supported me throughout my PhD. In particular, I want to thank **Helena Held**, **Axelle Bono**, **Marie Haenni**, **Marie Mayoly**, **Martha Nigg**, **Sarah Offner**, **Coralie Humair**, and **Cynthia Turci** for their unbreakable faith in me, for sharing their lives and their love with me even from so far, even after so much time.

I thank you, **Simon**. For your love that keeps me strong. For making one of the most challenging periods of my life also one of the brightest. For highlighting the beautiful parts of science to me even in my darkest moments. For believing in me. For teaching me so much, much more than any other scientist probably. For sharing with me this passion of knowledge, be it about physics, philosophy, psychology or biology, be it about people or things. You foster my understanding, push my reasoning to its limits, make of me a kinder, more modest and smarter person.

Finally I would like to thank my family. My parents, **Aldo** and **Monika**, who made all this possible by always believing in me without ever setting unwarranted limits to what I could achieve. My sisters and brothers, **Rebekka**, **Mauro**, **Diego** and **Rahel**, for having made any part of my life so easy and beautiful.

Contents

0	Summary	i
1	General introduction	1
1.1	Uncertainty	2
1.2	Of Probabilities and Statistics	3
1.2.1	Probabilities	3
1.2.2	Inferential statistics	5
1.3	Statistical reasoners	8
1.3.1	Are humans statistical reasoners?	10
1.3.2	Is statistical reasoning evolutionary ancient?	16
1.3.3	Open questions	17
1.4	Aims of PhD	18
2	Long-tailed macaques (<i>Macaca fascicularis</i>) can use simple heuristics but fail at drawing statistical inferences from populations to samples	21
3	Long-tailed macaques extract statistical information from repeated types of events to make rational decisions under uncertainty	23
4	Do infants and preschoolers quantify probabilities based on proportions?	25
4.1	Introduction	28
4.2	Methods	29
4.2.1	Participants	29
4.2.2	Materials	30
4.2.3	Procedure and design	30
4.2.4	Preference trial	31
4.2.5	Probability condition	32
4.2.6	Baseline condition	32
4.2.7	Coding	32

4.2.8	Predictions	33
4.3	Results	33
4.4	Discussion	34
5	General Discussion	38
5.1	Evolutionary origins of statistical reasoning	38
5.1.1	Reasoning about populations of objects and samples	38
5.1.2	Reasoning about repeated types of events	41
5.1.3	Differences between different types of reasoning	43
5.2	Of smart babies and dumb preschoolers	46
5.3	Future perspectives	47
5.3.1	Definitions	47
5.3.2	Considering mental representations	50
5.4	Conclusion	52
A	Supplementary material for: Do children and preschoolers quantify probabilities based on proportions?	I

List of Figures

4.1	Procedure in Placì, Fischer & Rakoczy (<i>submitted</i>)	31
4.2	Children's performance in Placì, Fischer & Rakoczy (<i>submitted</i>) . . .	34

Summary

Decision makers are often faced with different options and must make decisions without having perfect knowledge about the outcomes. They have only samples of data at hand, extracted from past and present states of their environment, that they must use to generalise and make predictions.

To help them in this process, humans developed in the 19th century an application of probability theory commonly referred to as *inferential statistics*, which deals with uncertainty inherent to inferences drawn from samples to populations. However, it is possible that humans were making statistical inferences way before the formalization of this branch of mathematics and that they possess, as part of their biological heritage, a set of probability intuitions that help them handle uncertain situations from infancy onward. It is also possible that such intuitions have deeper evolutionary roots and that humans' closest relatives, the nonhuman primates, also share such reasoning abilities.

These questions have fascinated scientists for several decades and have led to numerous studies with human adults, human children of various age categories, and recently, with nonhuman great apes. Findings suggested that human adults are statistical reasoners depending on how information is presented to them (e.g., natural frequencies vs. percentages). Research with human children yielded ambiguous results. Children from 5 years onward and infants seem to incorporate aspects of probability theory in their inferences. However, research with 3- and 4-years-olds suggests that they are devoid of probability intuitions. Research with nonhuman primates suggested that apes also share some probability intuitions that they use to make predictions from populations to samples.

The aim of my PhD thesis was twofold. First, I aimed to further investigate statistical reasoning abilities in nonhuman primates to shed more light on the evolutionary origins of this ability and to assess how different presentation formats affect nonhuman primates' reasoning abilities. Second, I aimed to shed more light on the development of probability intuitions during human childhood. In particular, I wanted to try and uncover why 3- and 4-years-olds were outsmarted by infants in the same kind of tasks.

In my first study, I tested whether long-tailed macaques, similar to human children and nonhuman apes, can make predictions from populations of preferred and dis-preferred food items to samples. I used different conditions to rule out that subjects rely on quantity heuristics (i.e., comparing absolute quantities of food)

rather than on proportions of food to make predictions. Results showed that long-tailed macaques, in contrast to children, apes, and capuchin monkeys (that were tested later on), do not rely on proportions but rather on heuristics to make inferences from populations to samples. These findings might be taken to suggest that the ability to draw this kind of inferences evolved in a convergent fashion in New World monkeys and apes. However, there are still open questions regarding why long-tailed macaques failed and regarding whether the other tested species really relied on proportions.

In my second study, I investigated whether long-tailed macaques can extract statistical information from repeated types of events to make rational decisions under uncertainty. Subjects were presented with different options associated with different probabilities of rewards. Subjects could interact several times with each option separately to learn the different probabilistic patterns of rewards. In a subsequent test trial, subjects could choose between both options presented simultaneously. Results revealed that long-tailed macaques extract statistical information from repeated types of events to make inferences about predictive factors and rational decisions. These findings suggest that, depending on how information is presented, long-tailed macaques are statistical reasoners.

In my last study, I tested whether human infants and preschoolers make inferences from populations of preferred and dis-preferred objects to samples, based on a paradigm already used with both age groups. In my study, however, all objects in the populations were visible to make sure that subjects could rely on the entire proportional information. Results showed that neither preschoolers nor infants relied on proportions to make inferences. However, it is not clear why infants failed, as they also failed in a baseline condition in which the outcomes were certain. One possibility is that infants had no strong preference for any object and therefore chose at chance. Another possibility is that the procedure confused them even if we kept it as close as possible to the procedure of previous studies. A last possibility is that infant's performance has been over-estimated in previous studies. In fact, it is possible that in previous studies, parents, who were not blind to the conditions, interfered with their child's decisions.

In conclusion, these findings suggest that statistical reasoning is evolutionary ancient, as different primates species (and bird species) engage in it. However, there might be more than one cognitive mechanism at play, as long-tailed macaques succeeded when statistical information could be experienced through repeated types of events but not when it had to be extracted from the present environment. These findings also suggest that the ability of infants to rely on proportions of objects

to make inferences about random sampling might have been over-estimated in the past. In general, more research is necessary to disentangle between different kinds of probability intuitions and their respective evolutionary origins and to understand how these intuitions develop during human ontogeny.

General introduction

“It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge”

- Pierre-Simon Laplace (1812)

“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?”

- Jean Piaget and Bärbel Inhelder (1975)

We (humans and other animals) do not know in what precise state our environment is, nor do we know how events will exactly unfold over time. We have to guess and base our decisions on these guesses. This is what statistical reasoning is about. Put simply, it is the ability to make predictions and rational decisions under uncertainty. But it is also much more, at least for us humans. It helps us in the process of studying and understanding the world in which we live. What exactly is uncertainty and how does one make predictions that lead to rational decisions under it? Is statistical reasoning a pure cultural invention of our species, or is it a cognitive adaptation for dealing with uncertainty that is present in humans from infancy onward, and that might be evolutionary older than us? In this dissertation, I will try to answer some aspects of these questions.

1.1 Uncertainty

Uncertainty is defined by a lack of information about how things were, are, and will be. If I stand at my open window and look outside, everything seems to be tinged with uncertainty. I consider the clouds rapidly moving in the sky and I wonder whether tomorrow will be sunny. I see the breeze gently rippling the lake and I ponder over the temperature of its waters. I cast an eye over the empty lawn and ask myself where my cats are.

Uncertainty becomes a problem when it comes to decision making. Every living organism capable of self-propelled movement needs to make decisions about where to go, the moment more than one option is available. Is it going to be left or right? Up or down? Similarly, every animal needs to select a food patch or a food element the moment several options are obtainable. River or field, fish or rodent? The same goes with partner choices or any other kind of choice. Uncertainty about what choice to make arises when the outcomes of each decision are not exactly known, either because there is no information about them at all, or because there is only partial information, in the present or from the past, that might or might not be useful for a new case. What is the safest option, right or left? Are there any fish in the river? Is this potential mate going to take good care of its offspring?

Uncertainty can be expressed in different degrees. Still standing at my window, I cannot say whether I will see the sun tomorrow, or whether the clouds will shade it from my view. However, not one second will I doubt that, clouds or no clouds, it will rise above the horizon. I do not know the exact temperature of the lake but I will definitely refrain from going swimming in it as I guess it is rather cold. My cats might not be visible, but I could bet a fair amount of money that they are in a ten kilometer radius of the house. Similarly, one can have different degrees of uncertainty regarding which decision would lead to higher benefits. Based on previous information, one might believe that the field is a better option at this time of the year, as the river is likely to be dried out or that this potential mate is probably going to disappear as soon as he passed on his genes while this other one will probably stay longer and invest in his offspring's upbringing.

Even under uncertainty there seems to be a way to differentiate between what is more or less probable. Or, in other words, even without knowing exactly what was, what is, and what will be, it seems possible to ascribe different probabilities to different states of the world and to different events, which, in turn, might encourage or hinder some decisions.

1.2 Of Probabilities and Statistics

What does it mean to ascribe a probability to an event? How can predictions be made and how can they direct decisions? To answer these questions, we have to enter the realm of probability theory, inferential statistics, and causality.

1.2.1 Probabilities

Concepts of probability have existed for a long time but the first evidence of their mathematical formalisation dates back only to the 16th century. Probability theory made its debut thanks to games of chance. Among its pioneers was the polymath Girolamo Cardano, who pondered over how mathematics could help solve problems linked to games of chance. But what is considered to have laid the foundation of probability theory is a correspondence exchanged in 1654 between the mathematicians Blaise Pascal and Pierre de Fermat.

In the following centuries, probability theory expanded, thanks to the collective work of many more eminent mathematicians, among whom were James Bernoulli, Abraham de Moivre, and Pierre-Simon Laplace (see Debnath & Basu, 2015, for more details on the history of probability theory). Laplace completed what is nowadays considered as the classical interpretation of probability. The concept of probability was discussed in examples that detached themselves from lotteries or games with one or several dice. It was applied to personal decisions, as well as to judicial and financial ones.

Probability theory had to wait until the 20th century to receive its set of axioms. These axioms were established by the mathematician Andrey Kolmogorov in his *Foundations of the Theory of Probability* (1933). The three axioms suggested by Kolmogorov are:

1. The probability that an event (denoted by A) occurs is equal to or greater than zero. It can be denoted by $P(A) \geq 0$.
2. The probability that at least one event belonging to the set of all possible events (denoted by E) will happen is 1. It can be denoted by $P(E) = 1$.
3. If two events are independent, then the probability that both events occur is equal to the sum of the individual probabilities of both events. It can be denoted by $P(A + B) = P(A) + P(B)$.

There are five different interpretations of probability that are considered as the leading views, and all of them respect Kolmogorov's axioms. These interpretations

are commonly referred to as the classical, the logical, the subjective, the frequentist, and the propensity interpretations. It is not in the scope of the present dissertation to discuss these views, their advantages and their limitations (see Hájek, 2012, for an elaboration on the different interpretations); however, I will briefly provide an overview.

According to the classical interpretation, the probability of an event happening depends on the total number of possible outcomes of equal probability, and on the proportion of these outcomes that is favourable to the event of interest. For example, to calculate the probability of obtaining a “6” when tossing a fair dice, one needs to consider the total number of outcomes that have equal chances of happening when tossing a fair dice, namely six (all the numbers from 1 to 6), and the number of possible outcomes favourable to our event of interest, namely one (the dice lands on “6”). Therefore, the probability of obtaining “6” is $\frac{1}{6}$.

According to the logical interpretation of probability, the probability of an event can also be measured by taking into account all the possible events that can happen, and by considering those that are in favour of the event of interest and those that are not. However, this time, the sample space of all possible events that can happen must not be made of events of equal probability. Instead, each event can be assigned a different weight depending on the evidence that logically supports it.

Proponents of the subjective interpretation consider probabilities as degrees of beliefs made by rational agents. Degrees of belief vary with the information (or priors) agents already possess and with the evidence. Probabilities as degrees of belief are updated according to Bayes’ theorem¹:

$$P(H | E) = \frac{P(E | H)P(H)}{P(E)}$$

In which H refers to the hypothesis whose truth needs to be assessed, and P(H) stands for the probability of that hypothesis being true, prior to the new evidence. E stands for the evidence. P(H| E), or posterior probability, stands for the probability that H is true given the evidence E. P(E| H) stands for the probability of observing E if H is true, and is called likelihood. P(E) is the marginal probability of the evidence, averaged over all possible hypotheses.

¹Note that Bayes’theorem is not only used to update degrees of beliefs. This theorem is directly derived from Kolmogorov’s axioms and is a formula to calculate a conditional probability. It can therefore be applied to all situations in which a probability must be calculated conditional on a given proposition (see Cosmides & Tooby, 1996, for an elaboration on the topic).

According to the frequentist interpretation, probabilities are long-term frequencies of events. For example, if one flips a coin one thousand times and it lands four hundreds times on "head", then the probability of getting a "head" by flipping that coin is 0.4. For frequentists, a probability is therefore independent from the human mind and from knowledge. It is a value anchored in the real world.

Proponents of the propensity interpretation regard probabilities as a description of the physical properties of things and of their interaction with the world. According to this view, if the face of a dice has a $\frac{1}{6}$ probability to occur, it is due to the physical properties of the dice itself. The difference between this view and the frequentist view is that for proponents of the propensity view, single-events can also be assigned a probability.

Despite this lack of consensus on what probabilities are, there is an agreement about how to handle them mathematically, as stated in Kolmogorov's mathematical definition of probability (Savage, 1972).

As any other branch of pure mathematics, probability theory provides a theoretical framework for the analysis of conceptual or abstract ideas. In his axioms, Kolmogorov only considers abstract or ideal events that are the result of random processes. By random processes Kolmogorov meant processes that produce sequences of events that cannot be expressed as a function of any shorter sequence but only as an identity function of itself or as a function of a sequence of the same length (but see Batanero, Green & Serrano, 1998, for different interpretations of randomness). In other words, random events are events that cannot be predicted based on previous information. Each new event is therefore independent from the previous one. The only known processes that seem to generate random sequences are found at the quantum level (Calude, Dinneen, Dumitrescu & Svozil, 2010).

But if so, what application does probability theory have in real life, in which a majority of events are considered precisely not to be random? One of the applications of probability theory is called inferential statistics.

1.2.2 Inferential statistics

Inferential statistics made its debut in the 19th Century with the work of Francis Galton and Karl Pearson (Varberg, 1963). It is a form of inductive inference (Savage, 1972), that is to say, it is concerned with drawing generalisations from limited data. There is uncertainty in this process because inferences go beyond the data given and there are many possible ways generalisations can be made. Inferential statistics is often described as the art of making inferences and decisions under uncertainty, as the study of uncertainty, or as the art of making inferences from samples to pop-

ulations (see for example Bandyopadhyay & Cherry, 2011; Lindley, 2000; Varberg, 1963).

There are several ways in which inferential statistics has recourse to probability theory. It is not in the scope of this dissertation to give an account of all these ways. I will only cover how inferential statistics works (for a detailed account, see Bandyopadhyay & Cherry, 2011). Because they are important for understanding what inferential statistics is, I will also introduce the concepts of statistical regularities, samples, and sampling processes.

Of statistical regularities

The world is not chaotic randomness, where everything and anything can happen with equal probability. If you look out of your window, you will not even consider the possibility that the sun might fall on the ground, transform itself into an elephant and start dancing the Macarena. Scientists believe that causality creates regularities with respect to how events follow each other over space and time. Sunsets follow sunrises in a regular manner. The same type of fruits appears in the same species of trees. Some individuals spend more time together than other individuals. Lung cancer develops more often in smokers than in non smokers, etc. These regularities can be expressed by relative frequencies of events and their covariation patterns with other events or with different variables (e.g., age, species, smoking habit, etc) and can be plotted as probability distributions (e.g., the probability of having lung cancer given different smoking habits).

Regularities can be weak or strong or anything in the middle. Pressing on the button of the remote control will almost always turn on the television, but a smoker will not necessarily contract lung cancer. Variability occurs when more than one causal factor affect a type of event in conjunction, but not all factors are always at work at the same point in space and time (see Cheng & Lu, 2017, for an elaboration on this issue). Imagine you follow for fifty years the lives of one hundred humans who smoke one packet of cigarettes every day. At the end of your survey, some humans will have contracted lung cancer and some not. This is because smoking alone is not sufficient to cause lung cancer but works in conjunction with other factors (that can be genetic or environmental) that were or were not at work in different individuals.

One of the aims of inferential statistics is to inform decision makers. Decision makers form hypotheses about the probability distribution of particular parameters and want to know whether they can act according to these hypotheses or whether they should modify or discard them. Consider a decision maker who wants to know whether the following hypothesis is true: the probability of contracting lung cancer

in humans increases with the frequency of smoking cigarettes. Ideally, she would collect information about the smoking habits of every human that has ever lived and about whether they contracted lung cancer. This of course is not possible. Our decision maker will have to rely on a sample of data, and she will need this sample to be representative of the population of humans.

Of samples and sampling processes

There are two important factors to consider for a sample to be representative of a population: sample size, and sampling process. If a sample is too small, it is unlikely that the measured regularities are representative of the real probability distribution. The law of large numbers, which comes from probability theory, states that given a random sampling process, the bigger the sample, the closer the value of the measured target parameter is to the target population parameter. (Bandyopadhyay & Forster, 2011). For a sample to be representative of a population, it needs to cover the whole set of possible outcomes (e.g., different degrees of smoking and whether they are associated with cancer or not) and the probability associated with each outcome. Imagine a small sample that by chance contains only data about non-smokers and heavy smokers. This sample would not be representative of the whole population of humans and would lead to imprecise predictions, because people who smoke one cigarette a month would be associated with the same probability of getting cancer than people who smoke two packets of cigarettes a day. Big samples are more likely to contain the whole set of possible outcomes and their respective probability than small samples.

The law of large numbers applies to randomly drawn samples. A randomly drawn sample is a sample that has the same probability as any other sample of being drawn from the population (Bandyopadhyay & Cherry, 2011). Why is this important? Imagine that your sample of humans, rather than being randomly selected from the population, included only construction workers manipulating material containing asbestos. In this sample, the probability distribution of having lung cancer given different smoking habits would be flattened, as the probability that non-smokers or occasional smokers have cancer would increase due to the deleterious effect of asbestos. If the information obtained from this sample was used to generalise to the whole population, erroneous predictions would be made. By selecting a sample randomly, alternative factors that can also affect a probability distribution are more likely to be present with the same frequency with which they are present in the population, especially if the sample is big.

Let us assume our decision maker obtained a large sample of data that were

randomly selected. She can now use different statistical methods to test her hypothesis that there is a correlation between increasing frequency of smoking and contracting lung cancer. She could, for example, calculate the probability with which the observed correlation in her sample, or a more extreme correlation, is sampled if the true correlation that holds in the population is in fact 0. If the probability she gets is lower than some criterion, she can decide to reject the null hypothesis and infer that the true correlation is greater than 0. She could also use Bayes' theorem and calculate the probability of her hypothesis being true, given her evidence, or she could build a mathematical model in which smoking is a factor, and one model in which smoking is not a factor, and compare which one predicts best the regularities found in her sample.

Once her assessment is done, our decision maker can use this new knowledge and decide whether she should stop smoking or not. Note that in my example the population of reference was a population of several objects (here humans), but a population can also only refer to one object. Let us assume a decision maker wants to assess whether a dice is fair. In this case, the population she should consider refers to all the possible outcomes that can be obtained by tossing this dice.

1.3 Statistical reasoners

In the previous section, we saw what probabilities are, what their mathematical definition is and how inferential statistics uses probability theory to inform decision makers. We saw that both probability theory and inferential statistics developed rather late in the history of humanity. However, this does not in any way mean that humans did not estimate probabilities and make statistical inferences before the 16th Century. A question that keeps fascinating scientist is whether the way humans reason under uncertainty in their everyday lives does incorporate aspects of probability theory. Are humans statistical reasoners or are inferential statistics and probability theory only products of human culture? In his *Théorie analytique des probabilités* (1812), Laplace wrote a sentence that is commonly translated as:

The theory of probabilities is at bottom nothing but common sense reduced to calculus; it enables us to appreciate with exactness that which accurate minds feel with a sort of instinct for which ofttimes they are unable to account.

And in fact, many mathematicians of the Enlightenment considered probability theory to be a description of human judgments (Daston, 1988).

In the following section, I will provide a short summary of the literature that investigated whether humans (both adults and children) are statistical reasoners and whether this ability might be evolutionary ancient and shared with other animals. Before this, however, it is important to get a more precise idea of what being a statistical reasoner implies.

Available definitions about what statistical reasoners should be able to do are rather general. Some examples are “reasoning under uncertainty” (see, for example, Denison & Xu, 2014), “making inferences that follow Bayes’ theorem” (see, for example, Gigerenzer & Hoffrage, 1995), or “making inferences from populations to samples and from samples to populations” (see, for example, Rakoczy et al., 2014; Xu & Garcia, 2008). There is also variability in the terms used to name the ability to reason under uncertainty. Some scientists speak of “probabilistic reasoning” (see, for example, Denison & Xu, 2014; Tecwyn, Denison, Messer & Buchsbaum, 2017), some others of “intuitive statistics” (see, for example, Eckert, Call, Hermes, Herrmann & Rakoczy, 2018; Rakoczy et al., 2014; Xu & Garcia, 2008), and others of “Bayesian reasoning” (see, for example, Gigerenzer & Hoffrage, 1995). All of these terminologies and definitions seem to imply the ability to use available information (past and/or present information) to make inferences under uncertainty that are congruent with probability theory. However, there is no indication about the range of probability problems statistical reasoners should be able to solve. If a reasoner complies to probability theory in some situations but not in others, do we call her a statistical reasoner nonetheless?

What is also not clear in the existing literature is what information statistical reasoners are expected to use. Inferential statistics, as an application of mathematics, is content-free: it can work with any information, be it social or physical, more or less abstract, represented by any kind of symbolic language, as long as it can be expressed numerically. What about statistical reasoning? In theory it should also be indifferent to the information it uses. In practice, however, it is limited by the brain in which it occurs. Each brain is limited by the information it can represent, be it an event, a hypothesis, or a factor (see, for example, Tenenbaum, Kemp, Griffiths & Goodman, 2011) and by the information it can integrate (Gigerenzer & Hoffrage, 1995; Marr, 1982), which are, in turn, limited by evolution (Cosmides & Tooby, 1996). In conclusion, there is rather a general feeling about what a statistical reasoner should be. However, available definitions still provide many degrees of freedom to researchers who want to investigate statistical reasoning.

1.3.1 Are humans statistical reasoners?

Adults

The literature about whether human adults are statistical reasoners is controversial. It seems that adults were statistical reasoners in the 1960s, lost their ability in the 1970s, and partially regained it in the 1990s.

In the 1960s, Ward Edwards and his colleagues (W. Edwards, 1968; Phillips & Edwards, 1966) developed the first experiments testing statistical reasoning in human adults. For example, in Phillips and Edwards (1966), subjects were asked to imagine 10 bags. A given number of these bags, say r , had a higher proportion of red chips and the rest a higher proportion of blue chips. Subjects were then shown a bag that the experimenter had supposedly randomly chosen out of the 10 possibilities. Subjects therefore had a prior probability of $\frac{r}{10}$ that this bag had a higher proportion of red chips and of $\frac{10-r}{10}$ that this bag had a higher proportion of blue chips. Subjects saw the experimenter draw chips out of this bag. Subjects had, after every draw, to indicate whether they believed that the bag had a higher proportion of red or of blue chips. Results revealed that subjects did update their belief about which bag the experimenter had drawn given the sampled chips. However they did so less than is predicted by Bayes' theorem. The conclusions of the authors were that adults are statistical reasoners, albeit conservative ones.

The 70s gave a new turn to the history of statistical reasoners, mainly due to the joint work of Daniel Kahneman and Amos Tversky (see, for example, Kahneman & Tversky, 1972, 1973), whose famous conclusion was that “[man] is not Bayesian at all” (Kahneman & Tversky, 1972, p. 450) and “people do not appear to follow the calculus of chance or the statistical theory of prediction” (Kahneman & Tversky, 1973, p. 237). From then on, humans adults seemed to be prey to all sorts of fallacies and to mainly rely on heuristics to make predictions and decisions under uncertainty. For example, subjects were given the following problem:

Problem 1: Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). 85% of the cabs in the city are Blue, and the remaining 15% are Green.

A cab was involved in a hit-and-run accident at night. A witness later identified the cab as a Green cab. The court tested the witness' ability to distinguish between Blue and Green cabs under nighttime visibility conditions. It found the witness able to identify each color correctly about 80% of the time, but confusing it with the other color about 20% of the time.

What do you think are the chances that the errant cab was indeed Green, as the witness claimed?

Results showed that most subjects neglected the base rate at which both types of cars operated in the city as well as the false-positive rate at which the witness identified the car as a Green cab. Subjects tended to answer 80%, which corresponded to the probability with which the witness correctly identifies a green car, while the correct answer was 41% (Bar-Hillel, 1980).

The view that humans rely on heuristics rather than on probabilities when making decisions under uncertainty predominated during the 80s but it started to be reexamined in the beginning of the 90s. The idea emerged that the information format that the brain takes in as input and transforms into an inference might play a role in how people perform in statistical reasoning tasks. New findings suggested that natural frequencies are easier to process than percentages (Cosmides & Tooby, 1996; Gigerenzer, 1996; Gigerenzer & Hoffrage, 1995, 1999). Other findings indicated that the representation of the problem rather than the numerical format is of importance. The representation of the problem can facilitate tasks by helping people imagine all the possibilities linked to this problem (Johnson-Laird, Legrenzi, Girotto, Legrenzi & Caverni, 1999).

What is not clear from this research is the role education plays in adults' statistical reasoning abilities. Subjects of these experiments were usually students who had learnt the basics of probability theory at school. Thus, these findings are not very informative about the natural state of humans' statistical reasoning abilities.

Children

Parallel to research with human adults, researchers also investigated intuitions about probabilities in human children. The second quote introducing this chapter was taken from a book by Jean Piaget and Bärbel Inhelder (1975), called "The origins of the idea of chance in children". This book summarises years of research with children, investigating whether humans are intuitive statistical reasoners. Piaget and Inhelder conducted many experiments to test different aspects of "intuitions of chance", such as intuitions about random mixture of objects, intuitions about outcomes of random drawings, and intuitions about combinatorics. In one experiment, for example, children were shown a bag containing red and blue marbles and had to predict the outcomes of random drawings. In the second part of the experiment, the bag was, unbeknownst to the children, replaced with a bag containing only blue marbles. The experimenter assessed the children's level of surprise when they discovered that only

blue marbles came out of this bag, and whether they understood that there was a trick or whether they considered it to be normal. The younger children expected the marbles being drawn out of the bag to follow some kind of order (for example, first only blue marbles, then only red ones, then blue marbles again), and they were not really surprised to see in the second part that only blue marbles were drawn. According to Piaget and Inhelder, this was due to a lack of experience and of intuitions about probabilities. Some of these young children explained this unlikely phenomenon by ascribing intentions to the balls (for example “the red ones don’t want to come out”, or “the blue ones are pushing the red ones”) and the majority did not conclude that the bag was only filled with blue marbles. Only older children seemed to update their priors about the proportions of red and blue marbles of the bag given the evidence. The main conclusion of Piaget and Inhelder was that children do not possess an intuition of probability from birth. Instead, they concluded that intuitions develop over time (beginning around 7 years of age) and depend on knowledge of logical and arithmetical operations.

However, several researchers criticized these experiments, saying that they were too difficult for children (see, for example, Braine, 1962; Yost, Siegel & Andrews, 1962). In fact, several studies showed that with tasks of reduced difficulty, children as young as 5-years-olds, as well as prenumerate adults, make inferences under uncertainty that are congruent with probability theory (Brainerd, 1981; Fontanari, Gonzalez, Vallortigara & Girotto, 2014; Pighin, Girotto & Tentori, 2017; Sobel, Tenenbaum & Gopnik, 2004; Yost et al., 1962).

Infants

In the past two decades, research on probability intuitions reached infants, that is, children around one year of age or less. Paradigms had to be adapted to be applicable to younger children, as younger children would not succeed in making explicit statistical inferences. Infants were therefore tested with looking-time, violation-of-expectation, and choice tasks. Findings from this research supported the idea that infants have intuitions about probabilities that they use to make inferences under uncertainty.

For example, it was shown that 12-month-olds ($N = 20$ for each experiment) look longer when objects of the minority type exit a virtual urn in which several objects of two different types randomly move (Téglás, Girotto, Gonzalez & Bonatti, 2007; Téglás et al., 2011), suggesting that very young children form expectations about the outcomes of random processes that agree with probability theory.

Others studies showed that infants use these intuitions to make inferences from

populations of objects to samples and from samples to populations. For example, in a study by Stephanie Denison and Fei Xu (2014), 12-month-olds ($N = 24$ for each experiment) were presented with two transparent buckets containing a mixture of black and pink lollipops. Before the actual trial, the experimenter assessed the preference of each child for one of the two lollipops. During the test trial, children saw the experimenter blindly draw a lollipop from each bucket and hide it in a cup so that only the stick of the lollipop was apparent. The experimenter then encouraged the child to crawl towards the cup of her choice. The content of each bucket was varied so that one bucket always contained a higher proportion of the favourite lollipops than the other bucket. Different conditions were tested. In some conditions, the absolute quantity of preferred lollipops was equal in both buckets, while in others it was higher in the unfavourable bucket. For example, in Experiment 2 there were 16 preferred and 4 non-preferred lollipops in the favourable bucket, and 24 preferred and 96 non-preferred ones in the unfavourable bucket. This composition of the ratios allowed to control for the possibility that subjects could have relied on the quantity heuristic: “choose the sample coming from the bucket with more preferred items”. Results revealed that a significant majority of children crawled towards the cup containing the sample drawn from the favourable population. This evidence suggests that infants relied on the proportions of both types of items in the buckets to make predictions about samples.

In two other studies (Denison, Reed & Xu, 2013; Xu & Garcia, 2008), 4.5-, 6- and 8-month-olds ($N = 16$ for each age group, except 8-month-olds, for which no sample size was provided) were first shown two boxes containing a mixture of red and white balls. In one box, there was a higher proportion of red balls and in the other a higher proportion of white balls. Children then saw the experimenter drawing six balls out of one of the two boxes, without knowing which one. The sample contained a majority of one type of balls. The content of the box was then revealed. Six- and 8-month-olds (but not 4.5-month-olds) looked longer when the proportion of the box did not match the proportion of the sample. These findings suggest that infants updated their two initial hypotheses about the content of the box in light of the new data. Additional findings obtained with the same paradigm showed that 11-month-olds vary their expectation if, prior to the sampling, the experimenter expressed a preference for a type of ball, and whether she looked inside the box while sampling, or whether she was blind (Xu & Denison, 2009).

There is also evidence suggesting that infants rely on the relationships between populations and samples of objects to learn about the properties of objects. For example, in a study by Hyowon Gweon, Joshua B. Tenenbaum, and Laura E. Schulz

(2010) 15-month-olds ($N = 15$ children per condition) were presented with a transparent box filled with two types of objects, blue and yellow balls. The box was always filled with 12 visible balls of one type, and 4 of the other type. The experimenter drew a number (either one, two or three depending on the condition) of blue balls out of the box, and each time, squeezed them to make them squeak. Subjects were then given a yellow ball, and the experimenter recorded how many times they tried to squeeze the yellow ball. Results were that children squeezed the yellow ball less frequently if the blue balls sampled out of the box were more likely to have been sampled intentionally. A plausible reason for why the blue balls were sampled intentionally is that only the blue balls could squeak. A sample was more likely to have been sampled intentionally if the likelihood that it was sampled randomly was low, as in the case in which three blue balls were sampled out of the population with a higher majority of yellow balls. These findings suggest that infants use outcomes of sampling processes to assess whether the processes were random or intentional. They then use this information to decide whether they can generalise the properties of objects to other objects or not. These findings not only suggest that infants rely on proportions of objects to make inferences about sampling events, but also that they use the content and the size of samples to make inferences about the sampling process.

Preschoolers

Infants seem to be intuitive statistical reasoners, 5-year-olds and older children as well. What about preschoolers (i.e., 3- and 4-years-olds)? Research on preschoolers suggest that they do not possess intuitions about probabilities. For example, in a study by Girotto and Gonzalez (2008), 3-year-olds, 4-year-olds, and older children (sample size very variable depending on age group and experiment) were presented with a bag in which three objects of the same type and one of a different type were hidden. Children had to predict which object type was more likely to be drawn out of the bag. Three- and 4-year-olds, but not older children, failed at making the correct prediction. Later on, Girotto and his colleagues (2016) applied the same paradigm used in Denison and Xu (2014) with two populations of preferred and dis-preferred items to 3- and 4-year-olds ($N = 48$). In contrast to infants, both 3- and 4-year-olds performed at chance level in these tasks.

Statistical reasoning to learn about the world

All the examples with children provided above investigated their ability to make inferences from populations of objects to samples and vice versa, and in most examples,

information about the populations was provided. In real life, however, inferences about the probability distributions of parameters within a population do not only apply to the distribution of objects of different types. They also apply to properties of objects or of individuals that are not perceivable and must therefore be induced from how these objects or individuals repeatedly interact with their environment.

Several findings suggest that young children rely on statistical regularities found in samples of events, to make inferences about properties of objects or of people. For example, in a study by Hyowon Gweon and Laura Schulz (2011), 16-month-olds ($N = 14$ per condition) saw one experimenter activate a green toy on several occasions. On every activation, the toy made some music. The child observed this while sitting on her parent's lap. Next to her, lying on a cloth, was a second red toy. Once the experimenter stopped playing with the toy, she either gave the same green toy or a yellow toy to the child. In both conditions, the toy did not make any music if the child tried to activate it. The child then could choose between asking help from her parent, or reaching towards the red toy. Results revealed that in the condition in which children were given the green toy, they turned to their parent for help, whereas in the other condition they tried to reach for the red toy. This evidence suggests that infants paid attention to the covariation pattern between who succeeded in activating the toy and the type of toy being played, to infer whether they were responsible for the toy not activating or whether the toy was not working. In the first condition, two agents (the experimenter and the child) tried to activate the same toy, so that the failure at making music covaried with the child. In the second condition, two agents tried to activate two different toys, so that the failure covaried with the child and the different toy.

In another study by Anna Waismeyer, Andrew N. Meltzoff and Alison Gopnik (2015), 24-month-olds ($N = 32$) observed an adult repeatedly place two different toys on a device, which then produced music or not. Children were then presented with both objects and could chose which one they wanted to place on the device. Results revealed that children took into account the relative frequency at which the placement of both objects on the device was followed by music in the observational phase (for example in Experiment 2, the placement of the first object co-occured with the production of music in 4 occasions out of 4, whereas the placement of the second object co-occured with the production of music only 4 times out of 12) to infer which object was more likely to activate the device.

In a study by Elisabeth S. Pasquini, Kathleen H. Corriveau, Melissa Koenig, and Paul L. Harris (2007), 3- and 4-years-olds (for 3-years-olds: $N = 21$; for 4-years-olds: $N = 20$) were first shown short films in which they saw two adults labelling

four familiar objects. One of the adults was always more accurate than the other, but the accuracy of both was varied in different conditions (i.e., 100% vs. 0%, 100% vs. 25%, 75% vs. 0%, and 75% vs. 25%). Next, children saw films in which both informants labelled the same unfamiliar object with conflicting names. Children were then asked to indicate the name of the unfamiliar object. Results showed that 4-year-olds trusted the more accurate informant in all conditions, whereas 3-year-olds trusted the more accurate informant only in the conditions in which she was 100% accurate. These results suggest that at least 4-year-olds rely on relative frequencies of past events to estimate the probability that an informant is providing reliable information.

In summary, it seems that from very early on, human children make inferences that are congruent with probability theory and that allow them to learn about the structure of their physical and social environment. This ability seems to have evolved independently from human culture as a powerful tool to deal with uncertainty.

1.3.2 Is statistical reasoning evolutionary ancient?

Findings about the statistical reasoning abilities of humans led to the following question: “is statistical reasoning evolutionary recent and only found in humans, or evolutionary ancient and shared with other animals?” (Rakoczy et al., 2014, p.61)

Research from the 1950s and the 1960s suggest that different animal species (e.g., species of cockroaches, bumblebees, fish, pigeons, rats and monkeys) are “probabilistic learners” (Bitterman, Wodinsky & Candland, 1958; Bullock & Bitterman, 1962; Longo, 1964; Meyer, 1960; Real, 1991; Wilson, Oscar Jr & Bitterman, 1964). In these studies, individuals were typically presented with two different options on repeated occasions (usually several hundreds or thousands of trials). One option was leading to more rewards than the other. Results of these studies indicated that several species of animals rely on samples of rewarding patterns to direct their choices under uncertainty. However, the absolute and relative frequencies of rewards associated to each option were usually confounded in these experiments, making it impossible to conclude that individuals estimated probabilities.

During the 1980s and afterwards, many studies showed that animals make optimal decisions by considering prior probability distributions and new observations in different contexts such as foraging or mate choice (see for example Lima, 1984, 1985, Biernaskie, Walker & Gegeer, 2009; for a review see McNamara, Green & Olsson, 2006, Valone, 2006). However, these studies were based on behavioural observations rather than on experiments that systematically tested statistical reasoning and did not rule out alternative hypotheses. Also, it is hard to compare

optimal foraging behaviour to the behaviour previously described in human infants. Are the same cognitive mechanisms involved in all species? Is animal decision making under uncertainty also a flexible and domain-general ability that can be used in never encountered scenarios?

To bring some answers to these question, Hannes Rakoczy and colleagues (2014) investigated statistical reasoning in apes using a paradigm initially developed for human infants (Denison & Xu, 2014). Subjects were repeatedly presented with two transparent buckets containing a mixture of two types of food, a preferred type and a neutral type. The experimenter simultaneously drew one item out of each bucket while looking up and presented both fists to the subjects. By pointing towards one of the two hands, subjects received the item hidden inside. To maximize the probability to receive a preferred food item, subjects had to rely on the different proportions of food contained in the buckets. Different experiments were used to rule out that subjects relied on a comparison of absolute quantities rather than of proportions to make predictions. Overall results from these experiments revealed that apes, similar to human infants, preferred the sample drawn out of the population with a higher proportion of preferred food items. These findings suggest that some aspects of statistical reasoning are evolutionary ancient and shared at least with our closest relatives.

Capuchin monkeys were also tested in a paradigm in which they had to make inferences about the causal properties of two different objects based on information sampled from repeated interactions between these objects and a device they could or not activate (B. J. Edwards et al., 2014). Findings suggest that subjects relied on the frequencies at which the different objects activated the device, together or separately, to predict which object was more likely to activate the device on its own. However, this study did not systematically test whether subjects relied on relative frequencies of activation to make inferences, rather than on absolute frequencies. It is therefore not clear whether capuchins estimated probabilities and whether non-human primates in general extract statistical information from repeated types of events to make inferences under uncertainty.

1.3.3 Open questions

When I started my PhD, an increasing number of findings suggested that children were smart little scientists who formed theories of the world and updated them following the laws of probability (Xu & Kushnir, 2013; Xu, Kushnir & Benson, 2012). Additional findings suggested that intuitions about probabilities were evolutionary ancient and shared with at least our closest relatives the great apes.

Nonetheless, there were still many open questions in the literature of statistical reasoning. For example, it seemed that apes shared with human children the ability to make inferences from populations of objects to samples but what about other primates? Are all primates statistical reasoners because their last common ancestor already engaged in statistical reasoning, or is statistical reasoning more recent and evolved only in the apes' lineage? And what about statistical reasoning that is not constrained to inferences about populations and samples of objects? There was for example not much information about how nonhuman primates reason about statistical information extracted from repeated types of events. However, this ability is crucial to learn about the hidden properties of different social and physical environments.

There were also still open questions regarding the development of statistical reasoning during human childhood. More particularly, why were 3- and 4-year-olds outsmarted by 1-year-olds? Different explanations could account for these findings. The first one is that the ability of infants was over-estimated and that statistical reasoning, as concluded by Piaget and Inhelder (Piaget & Inhelder, 1975), only develops later during childhood. It is possible that infants succeeded in different tasks by relying on simpler strategies and statistical reasoning occurs only around 5 years of age. A second explanation could be that the ability of 3- and 4-year-olds have been under-estimated. They were maybe not motivated enough by the possible rewards. A third explanation is that these findings show the real abilities of children and statistical reasoning follows a U-shaped curve during development. More information was necessary to disentangle between these possibilities.

1.4 Aims of PhD

My PhD had two main goals. The first goal was to shed more light on the evolutionary origins of statistical reasoning. For this, I developed two studies testing the statistical reasoning abilities of long-tailed macaques (*Macaca fascicularis*), a species of Old World monkeys. The shared ancestor between humans and Old World monkeys dates further back than the ancestor shared between humans and nonhuman apes. If Old World monkeys also possess statistical reasoning abilities, it could indicate that their last common ancestor with apes possessed this ability as well.

My first study (Placi, Eckert, Rakoczy & Fischer, 2018) was based on a paradigm initially developed for human children (Denison & Xu, 2014), and later adapted to nonhuman primates (Rakoczy et al., 2014). I aimed to test whether long-tailed macaques make inferences from populations of objects to samples by relying

on the proportions of objects in the populations to estimate probabilities of different outcomes. Monkeys were presented with populations of preferred vs. neutral food items and had to rely on the proportions of the different food types in order to increase their chances to choose the sample containing a preferred food item. As in the previous studies, several conditions were tested to rule out that subjects relied on quantity heuristics.

In my second study (Placì, Padberg, Rakoczy & Fischer, 2019), I investigated whether long-tailed macaques rely on limited data extracted from repeated types of events to make rational decisions under uncertainty in more or less complex scenarios. Monkeys were, in a sampling phase, repeatedly presented with different options associated with different rewarding patterns. In a subsequent test trial, monkeys could choose between the two options presented simultaneously. I deconfounded absolute and relative frequencies of rewards to disentangle whether monkeys relied on quantity heuristics or on probabilities. The nature of the different options was also varied to explore the boundaries of this ability. For example, in the first experiment the two options were two different objects associated with different rewarding frequencies, whereas in the second and third experiment, the options were two humans who were more or less good at opening one or two different boxes containing rewards. The third experiment was even more complex. In one condition, for example, one human was better at opening both boxes than the other human, but both boxes had the same likelihood to be opened; if monkeys only focused on the number of food items delivered from both boxes, they would not notice this difference and would not have a preference for any option. Similarly, in the second condition, one box delivered more rewards than the other, but both humans opened boxes at the same frequency. Again, if monkeys only focused on which human delivered more rewards, they would not succeed in this task.

The second aim of my PhD was to uncover the factors that allow infants to succeed in tasks in which preschoolers failed. Shedding more light on this could help understand better how statistical reasoning develops over age, and/or help improve methods that test statistical reasoning. In my third study (Placì, Fischer & Rakoczy, *submitted*), I therefore investigated whether the abilities of 12-month-olds had been overestimated or whether the ability of 3- and 4-year-olds had been underestimated. For this I developed a paradigm based on Denison and Xu (2014) in which children were again presented with two populations of preferred vs. dis-preferred items and had to make predictions about samples drawn out of these populations. This time, however, all items in the populations were visible, so as to make sure that children did not rely on the quantity heuristic: “select the sample coming from the population

with more visible preferred items” (see Placi et al., 2018; Téglás, Ibanez-Lillo, Costa & Bonatti, 2015, and Chapter 4 of this dissertation), as this heuristic was not ruled out in previous studies. I also tried to create a strong preference for one item to ensure that children remained motivated to solve the tasks. These three studies are the focus of Chapter 2, 3 and 4.

Long-tailed macaques (*Macaca fascicularis*) can use simple heuristics but fail at drawing statistical inferences from populations to samples

**Long-tailed macaques (*Macaca fascicularis*) can
use simple heuristics but fail at drawing statistical
inferences from populations to samples**

Sarah Placi^{1,2,3}, Johanna Eckert^{2,3,4}, Hannes Rakoczy^{2,3}, Julia Fischer^{1,3}

¹Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

²Department of Developmental Psychology, University of Göttingen, Göttingen, Germany

³Leibniz ScienceCampus Primate Cognition, German Primate Center, Göttingen,
Germany

⁴Department of Developmental and Comparative Psychology, Max Planck Institute for
Evolutionary Anthropology, Leipzig, Germany

Long-tailed macaques extract statistical information from repeated types of events to make rational decisions under uncertainty

**Long-tailed macaques extract statistical information
from repeated types of events to make rational
decisions under uncertainty**

Sarah Placi^{1,2,3}, Marie Padberg^{1,3}, Hannes Rakoczy^{2,3}, Julia Fischer^{1,3}

¹Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

²Department of Developmental Psychology, University of Göttingen, Göttingen, Germany

³Leibniz ScienceCampus Primate Cognition, German Primate Center, Göttingen,
Germany

Scientific Reports 9(12107) (2019)

doi: 10.1038/s41598-019-48543-0

Chapter 4

Do infants and preschoolers quantify probabilities based on proportions?

Do infants and preschoolers quantify probabilities based on proportions?

Sarah Placì^{1,2,3}, Julia Fischer^{1,3}, Hannes Rakoczy^{2,3}

¹Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

²Department of Developmental Psychology, University of Göttingen, Göttingen, Germany

³Leibniz ScienceCampus Primate Cognition, German Primate Center, Göttingen,
Germany

Submitted

Abstract

Most statistical problems encountered throughout life require the ability to quantify probabilities based on proportions. Recent findings on the early ontogeny of this ability have been mixed: For example, when presented with jars containing preferred and less preferred items, 12-month-olds, but not 3- and 4-years-olds, seem to rely on the proportions of objects in the jars to predict the content of samples randomly drawn out of them. Given these contrasting findings, it remains unclear what the probabilistic reasoning abilities of young children are and how they develop. In our study, we addressed this question and tested, with identical methods across age groups and similar methods to previous studies, whether 12-months-olds and 3- and 4-years-olds rely on proportions of objects to estimate probabilities of random sampling events. Results revealed that neither infants nor preschoolers do. While preschoolers' performance is in line with previous findings, infants' performance is difficult to interpret given their failure in a control condition in which the outcomes happened with certainty rather than a graded probability. More systematic studies are needed to explain why infants succeeded in a previous study but failed in our study.

Keywords: intuitive statistics, probabilistic reasoning, developmental psychology, intuitive statistics; child cognition

4.1 Introduction

Probability judgments are an important component of rationality as they help people learn and make decisions under uncertainty. An increasing number of findings suggests that young children have intuitions about probabilities that help them make predictions about outcomes of random events (Téglás et al., 2007, 2015, 2011). In addition, these intuitions seem, under some conditions, to help them make rational decisions under uncertainty (Denison & Xu, 2014). However, within these findings, there are some contradictions about how children quantify probabilities and about how these intuitions develop during childhood. In particular, it is not clear whether young children are able to quantify probabilities based on proportions, an ability necessary to most probability problems (Bryant & Nunes, 2012).

In a study with 12-month-olds, for example, infants were first shown a pink and a black lollipop (Denison & Xu, 2014). The experimenter placed both lollipops on the ground and waited for the infant to crawl to one of them. The item chosen by the child was considered as her preferred item, and the other item as the less preferred one. The experimenter then presented two transparent jars containing a mixture of both types of lollipops to the child. In one jar, there was a higher proportion of preferred lollipops than in the other jar. The experimenter also brought two opaque cups that she placed on each side of the jars. She then concealed the content of both jars with two paper occluders, closed her eyes, reached into the jar on the right, grabbed a lollipop, and while holding it concealed in her fist so that the stick was the only thing apparent, she moved it to one of the two cups. She then repeated the same action with the jar on the left. The experimenter then encouraged the child to come choose one of the two cups. Several conditions were tested: in some conditions, the absolute quantity of preferred lollipops was equal in both jars, and in others, it was higher in the unfavourable jar (i.e. the jar with a higher proportion of less preferred lollipops). For example, in Experiment 2, there were 16 preferred and 4 less preferred lollipops in the favourable jar, and 24 preferred and 96 less preferred lollipops in the unfavourable jar. This was important to rule out that children solved the task by comparing absolute quantities of preferred items in both jars instead of estimating probabilities based on proportions. Results revealed that more children than expected by chance moved towards the cup containing the sample drawn out of the favourable jar and suggest that infants rely on proportions of large quantities of objects to make inferences about random drawing.

In three similar tasks (Giroto et al., 2016)), one of which was tested with 3-, 4- and 5-year-olds (Study 1), and the two others only with 3-year-olds (Study 2 and 3), 3- and 4-year-olds consistently failed at quantifying probabilities based on

proportions. Currently, we do not know why 12-month-olds performed better than 3- and 4-years-olds. The main differences between the lollipop study and Study 2 in Girotto et al. (2016) – which was the most similar in methods to the lollipop study – were (1) the use of lollipops in one case and of spoons with stickers in the other, (2) the difference in age between the subjects of both studies, (3) the fact that in the lollipop study, parents were not blind to the experiment.

One possible explanation could be that the older children were not motivated enough by the stickers and did therefore not care enough to solve the task. Another explanation is that the younger children relied on an alternative strategy to solve the task. For example, infants could have had the illusion that there were always more preferred lollipops in the favourable jar and could have solved the task by comparing absolute quantities of visible preferred items. Consider Experiment 2 again. In the unfavourable jar there were more preferred lollipop than in the favourable jar, however, it is possible that most of the preferred lollipops of the unfavourable jar ($n = 24$) were hidden by the high quantity of less preferred lollipops ($n = 96$) so that in appearance more preferred lollipops were seen in the favourable jar. A third explanation is that infants, instead of solving the task by estimating probabilities, simply responded to slight cues given by their parents who were not blind to the conditions.

The current study aimed to shed light on these unsolved questions. Our aim was to test for the ability to quantify probabilities based on proportions across two age groups using the exact same methods. We tested 12-month-olds and 3- and 4-years-olds in a paradigm similar to the one used in Denison & Xu (2014) and Girotto et al. (2016) to which we made some changes. In our study, the objects were not stacked one on top of the other in their respective containers. We selected big enough boxes so that each item could be seen separately and so that children could not solve the task by comparing absolute quantities of objects. We also changed the nature of the objects to make them more interesting as potential rewards. Finally, in our study, parents were blind to all conditions.

4.2 Methods

4.2.1 Participants

Sixty-eight children were included in the final sample, which was split into two age groups: infants ($n = 34$, mean age = 12 months, age range = 10-14 months) and preschoolers ($n = 34$, mean age = 46 months, age range = 36-54 months). Thirty-five additional children (thirty-four infants and one preschooler) were tested but not

kept in the final sample because they either did not make any choice in the preference test or the test condition ($n = 6$), or because they did not make the expected choice in the preference trial ($n = 29$). Children were recruited from a database of families who had voluntarily registered and agreed to participate. They were from mixed socioeconomic backgrounds and some family were plurilingual.

4.2.2 Materials

Objects

Transparent white and transparent blue Kinder eggs were used as stimuli for the experiment. To all eggs were attached thin wooden sticks of approx. 20cm in length, that allowed children to see that only one object was removed from each box and transferred to a cup. On all blue eggs were also stuck white cartoon eyes to make them more interesting. In addition, all blue eggs contained finger puppets representing different animals whereas all white eggs were empty (see Fig. S1 of Appendix C).

Population boxes

Two cubical transparent boxes (length \times width \times height: 37,8 cm \times 39,6 cm \times 18,5 cm) were used as container for the populations of Kinder eggs. Two white rectangular blankets were used to cover the boxes.

Cups

Two opaque cups (length \times width \times height: 10 cm \times 10 cm \times 15 cm) were used as containers for the eggs sampled from the boxes. Each cup was covered by a black tissue with a hole to allow the experimenter to hide the eggs inside.

4.2.3 Procedure and design

Each child was tested individually in a quiet room in presence of a family member. Children were told they would participate in a game. They sat on the lap of one parent (except for some 3- and 4-years-olds who preferred to sit alone, with the parent sitting next to them) who sat on a chair in front of a table. The experimenter sat on the opposite side of the table. The two cups were already placed on the two extremities of the table on each side of the child and the experimenter. Parents were asked to wear opaque glasses that prevented them from seeing the experiment. They were allowed to withdraw the glasses in between each phase of the experiment.

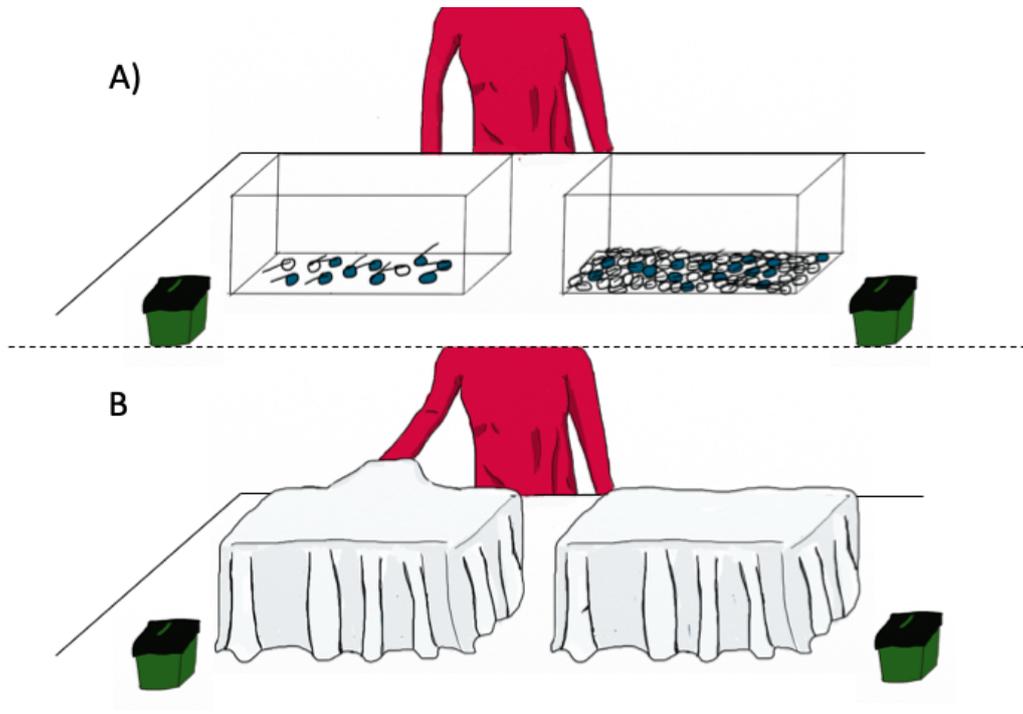


Figure 4.1: Procedure. The experimenter presented both boxes of the Probability condition, one after the other, as depicted in A). She then covered both boxes with white blankets and while closing her eyes, removed one object in the right box and hid it in the green cup adjacent to the box, as shown in B). The same was repeated with the second box. She then took both cups, approached them at equal distance from the child and encourage the child to choose a cup. The same procedure was repeated with the Baseline condition.

Each child would undergo a preference test, followed by a probability condition and a baseline condition (within-subject design).

4.2.4 Preference trial

The experimenter brought from under the table a transparent white egg and, drawing the attention of the children to it, said “Look here, I have a white egg, I open it. Oh, it is unfortunately empty. See, there is nothing inside”. The experimenter closed the egg and then brought a blue egg from under the table and said “Look here, I also have a blue egg, I open it. Oh, there is something inside”. She closed the blue egg again. Then, always starting with the white egg, she said “Look here” and slowly hid it in one of the cups, then repeated the same with the blue egg. She finally lifted both cups simultaneously and presented them at equal distance from the child who could try and grab the one she preferred. If the child chose the cup containing the blue egg, the experimenter said: “Oh great, you found the blue egg”. She opened the egg and allowed the child to play with the finger puppet for some

time. If the child chose the cup containing the white egg the experimenter said: “Oh, this egg is unfortunately empty”. Only children who chose the cup containing the blue egg were kept in the final sample.

4.2.5 Probability condition

The experimenter brought the two covered boxes on the table. One box contained 9 blue eggs and 3 white eggs (favourable population), the other box contained 16 blue eggs and 48 white eggs (unfavourable population). The experimenter uncovered the first box, always starting with the one on the right (see Figure 4.1). She said “Look what is in this box” and approached the box towards the child to allow her to clearly see its content. The experimenter shook the box to make clear that the eggs could be mixed. The experimenter repeated the same action with the second box. She then covered the two boxes again and, with her eyes closed, reached in the box on the right, hid one egg in her hand and placed it in the cup adjacent to the box. She repeated the same action with the second box. She finally lifted both cups simultaneously and presented them at equal distance from the child, saying: “pick the one you want” and waiting for the child to try and grab one cup. If the child chose the cup containing the blue egg the experimenter said: “Oh great, you found the blue egg”. She opened the egg, and allowed the child to play with the finger puppet for some time. If the child chose the cup containing the white egg the experimenter said: “Oh, this egg is unfortunately empty”.

4.2.6 Baseline condition

The same procedure as in the probability condition was repeated, only this time one box contained 48 blue eggs (favourable population) and the other box 48 white eggs (unfavourable population). This condition was added to check whether children understood the procedure. If children failed in this condition, it could not be due to a lack of intuitions about probabilities, as the outcome of the drawing was certain. The side of the favourable sample and of the favourable population was counterbalanced between the preference trial, the probability condition and the baseline condition.

4.2.7 Coding

Every session was video recorded. Children’s choices were coded from the videos. Whenever children chose the sample stemming from the favourable population, we considered it as a success and as a failure when they chose the alternative option.

4.2.8 Predictions

In the probability condition, we predicted that if children rely on a comparison of absolute quantities of preferred items to make predictions, they would choose the sample drawn out of the unfavourable population at above chance level as at this age, they can discriminate 2:3 ratios (Xu & Arriaga, 2007) and should therefore be able to distinguish that there are more blue eggs in the unfavourable population. If they rely on probabilities, we predicted that children would choose the sample drawn out of the favourable population at above chance level.

4.3 Results

In the baseline condition, 18 out of 31 (58%) infants selected the cup with the sample drawn out of the favourable population, no different from chance (binomial test, $p = .47$; 95% Confidence interval [39,75]). Twenty-four out of 34 (71%) preschoolers selected the cup with the sample drawn out of the favourable population, significantly different from chance (binomial test, $p = .02$; 95% Confidence interval [53,85], see Figure 4.2).

In the probability condition, 17 out of 34 (50%) infants selected the cup with the sample drawn out of the favourable population, no different from chance (binomial test, $p = 1$; 95% Confidence interval [32,68]). Eighteen out of 34 (53%) preschoolers selected the cup with the sample drawn out of the favourable population, no different from chance (binomial test, $p = .86$; 95% Confidence interval [35,70], see Figure 4.2).

Among the preschoolers who chose the right sample in the baseline condition, 58% (14 out of 24) also succeeded in the probability condition, no different from chance (binomial test, $p = .54$; 95% Confidence interval [37,78]).

To test whether age had an effect on infants' and preschoolers' performance, we did, for each condition and each age category, a post-hoc logistic regression, with subjects' choice as response variable and age (in month) as predictor. For this we used the glm function of the R package lme4. Results showed that age had no effect in any age category and any condition (Infants in probability condition: Estimate \pm SE = 0.39 ± 0.25 , $z = 1.52$, $p = .13$; Preschoolers in probability condition: Estimate \pm SE = -0.05 ± 0.07 , $z = -0.70$, $p = .49$; Infants in baseline condition: Estimate \pm SE = 0.23 ± 0.27 , $z = 0.85$, $p = .39$; Preschoolers in baseline condition: Estimate \pm SE = -0.05 ± 0.08 , $z = -0.70$, $p = .48$)

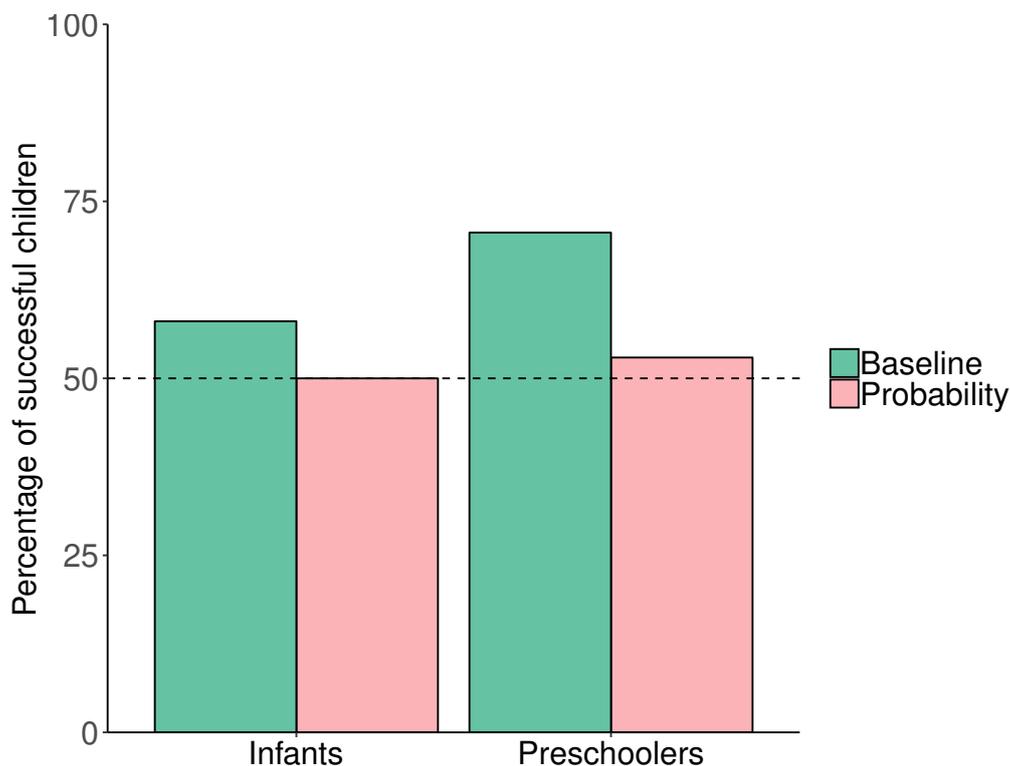


Figure 4.2: Percentage of successful children in each condition and each age category.

4.4 Discussion

In summary, we found that in our study, neither 12-months-olds nor 3- and 4-years-olds selected the favourable sample above chance level in the probability condition, and only 3- and 4-years-olds did so in the baseline condition. These results suggest that preschoolers did not rely on proportions of objects in the boxes to make inferences about the samples in the probability condition. We can rule out that their failure was due to the procedure, as in the baseline condition they selected the correct cup at above chance level. These results replicate the results of Girotto et al. (2016). It is nonetheless intriguing that in the baseline condition only 71% of preschoolers selected the correct cup. This cannot be attributed to a weak preference for the blue eggs, as only one preschooler out of 35 did not choose the cup containing the blue egg in the preference test (and was therefore not included in the final sample). It seems that something in the procedure confused some children. This confusing factor could have been the experimenter performing the drawing. In Study 3 of Girotto et al. (2016), the drawing was performed by a machine and children's performance in a condition similar to our baseline condition was better (90% success) than in Study 2 of Girotto et al. (2016; 69% success) and better than in our baseline condition, in which the experimenter did the drawing. Maybe some

children expected some sort of trick to be played on them and did therefore not consider the outcome of the drawing to be certain.

One could wonder whether there was a subsample of children who were not confused by the procedure and who performed well in the probability condition. This does not seem to have been the case, as only 58% of the children who performed well in the baseline condition also performed well in the probability condition.

Our findings also indicate that infants did not rely on the proportions of objects in the boxes to make inferences about the samples in the probability condition (i.e., did not preferentially choose the sample drawn out of the box with more preferred items). However, this is not surprising given that infants in this procedure seem not to have relied on any numerical information at all as indicated in the fact that they also performed at chance level in the baseline condition. A possible explanation for infants' failure in both conditions is that they did not have a strong preference for the blue eggs. In fact, almost half of the infants tested in the preference trial (28 infants out of 62) did not select the cup containing the blue egg and were therefore excluded from the final sample. Maybe, rather than having selected a sample of children with a preference for the blue eggs, we selected a sample of children who chose randomly between both egg colours. It is hard to explain why infants would not have a strong preference for an egg with cartoon eyes that visibly contains something when the alternative is an empty egg without eyes. In Denison & Xu (2014), it is not clear whether infants had a strong preference for any lollipop, as the only indicator for such preference is their good performance in the test trials. With only one preference trial it is in fact difficult to distinguish between a preference for an object and a random choice, especially if half of the subjects select one object and half select the alternative. It is also worth noting that there was a difference between our preference trial and the one in Denison & Xu (2014). In our preference trial, we placed the objects in the cups instead of placing them visibly in front of the children. Children therefore had to keep track of the location of both eggs. However, as infants are able to retrieve hidden objects from 8-months onward (Munakata, McClelland, Johnson & Siegler, 1997) and to find different objects in different locations (Feigenson & Halberda, 2004), this should not explain why only half of the infants in our preference trial chose the cup containing the blue egg.

Another difference between our experiment and the experiments in Denison & Xu (2014) is that in our case, parents could not have influenced their child's behaviour, as they were blind to the experiment. In Denison & Xu (2014) parents were not blind. Even if parents were asked not to interfere with the task, they might have done so inadvertently. It is therefore still not possible to distinguish whether

infants performed well in Denison & Xu (2014) because they received help from their parents, or whether infants performed poorly in our study because they did not have any preference for any object. More studies are necessary to disentangle between these possibilities. Future studies should assess children's preferences with several preference trials and should make sure that parents cannot interfere with their child's decisions. It is also worth asking whether this paradigm does need any other further improvement, considering the all-but-perfect performance of children in the baseline condition. It seems that there is something in the procedure that adds noise to the data and that is independent of children's ability to estimate probabilities.

In summary, the present study failed to find evidence that preschoolers rely on proportions of objects to make inferences about sampling events. These findings are congruent with previous results (Giroto et al., 2016). The present study also failed to find evidence for intuitive statistical inferences in infants, but these findings remain difficult to interpret in light of infants' failure in the non-probabilistic baseline condition. Our study also highlights the importance of improving this paradigm, both for its use with infants and older children, if we want to come to a better understanding of children's probabilistic reasoning abilities.

Acknowledgments

We thank Marlen Kaufmann, Isabel Ganter and Lydia Schidelko for their help with the data collection, and all parents and children who participated in our study.

Ethical statement

This research was conducted in accordance with the Declaration of Helsinki and the Ethical Principles of the German Psychological Society (DGPs), the Association of German Professional Psychologists (BDP), and the American Psychological Association (APA). It involved no invasive or otherwise ethically problematic techniques and no deception (and therefore, according to National jurisdiction, did not require a separate vote by a local Institutional Review Board; see the regulations on freedom of research in the German Constitution (§ 5 (3)), and the German University Law (§ 22)).

Funding statement

This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – Project number 254142454 / GRK 207.

Data accessibility

Supporting material has been uploaded as part of the electronic supplementary material. The datasets generated and analysed during the current study, as well as the R script used for the analyses, have been deposited with Dryad and can be accessed at (<https://datadryad.org/stash/share/c1Ur4Eo5bfx6uPsw69kKgeS6m8LJK6b9oYd4zpdep8>)

Competing interests

The authors have no competing interests

Authors' contributions

All authors contributed to the conception and the design of the study, SP collected, analysed and interpreted the data, SP drafted the manuscript. JF and HR commented on the manuscript and edited the text.

General Discussion

The aims of my PhD were to investigate the evolutionary origins of statistical reasoning and to shed more light on the developmental trajectory of intuitions about probabilities in human children. Based on the results of my studies and on results from other studies, including studies published after I started my PhD, I will discuss what these findings tell us about the two main questions of my dissertation. I will also discuss potential improvements that could be made in the field of comparative and developmental psychology to further investigate statistical reasoning.

5.1 Evolutionary origins of statistical reasoning

5.1.1 Reasoning about populations of objects and samples

In my first study, I investigated whether long-tailed macaques rely on cues of random sampling and on proportions of objects to make inferences from populations to samples. I included different conditions to rule out that subjects did not rely on quantity heuristics. Results were that long-tailed macaques, unlike human infants (Denison & Xu, 2014), apes (Eckert, Call et al., 2018; Rakoczy et al., 2014), capuchin monkeys (Tecwyn et al., 2017), and one parrot (Clements, Gray, Gross & Pepperberg, 2018), failed at making inferences from populations to samples that were congruent with probability theory. These findings suggest that statistical reasoning might have evolved in convergent fashion in New World monkeys, apes and birds. Before drawing hasty conclusions about the evolutionary origins of statistical reasoning, however, we need to clarify some points.

The first point is whether long-tailed macaques' poor performance signifies that they lack intuitions about probabilities. It does not necessarily. There might be a difference between long-tailed macaques and the other tested species in how they solved these tasks, as long-tailed macaques performed worse than the other

species. However, it needs not be a difference in statistical reasoning abilities. To be successful in the different tasks presented in Denison and Xu (2014), in Rakoczy et al. (2014), in Tecwyn et al. (2017), and in Placì et al. (2018), individuals had to be able to represent the sample space, that is, the food items contained in the population and their respective probability of being drawn. The actual proportions of food were only informative under the assumption that all items in the population could be sampled with the same probability. This in turn, should only be assumed if the sampling was random. This assumption should have been made based on the behaviour of the experimenter who closed her eyes or looked up while selecting items in the populations of food, and who therefore was blind to what she was drawing – a good indication for unbiased sampling. Long-tailed macaques could have failed at representing the proportions of food as the sample space because of the high salience of food. Using food as numerical information and as reward does in fact lower the performance of long-tailed macaques in quantity discrimination tasks (Schmitt & Fischer, 2011). However, this is also the case for human children (Carlson, Davis & Leach, 2005) and apes (Boysen & Berntson, 1995), so this explanation alone cannot account for why long-tailed macaques performed worse than the other species. Long-tailed macaques could also have failed in these tasks because they did not consider the experimenter to be blind to the sampling, either due to a lack of perspective taking – at which long-tailed macaques tend to fail, at least with humans partners (Kummer, Anzenberger & Hemelrijk, 1996) – or to a belief that the experimenter haptically distinguish between both food types, which was actually true. Moreover, the fact that two individuals within the group of long-tailed macaques were consistently good across the different conditions suggests that solving these tasks is not out of reach for animals of this species.

A second point that needs to be addressed is what a good performance in these tasks actually implies. Does a good performance clearly indicate inferences that are congruent with probability theory? If all conditions necessary to rule out reliance on different quantity heuristics are tested, a good performance across conditions would be a good indicator of intuitions about probabilities. This was, however, neither the case in studies with human infants (Denison & Xu, 2014), apes (Eckert, Call et al., 2018; Rakoczy et al., 2014), and capuchin monkeys (Tecwyn et al., 2017) nor in the study with the parrot (Clements et al., 2018). In the capuchin study, individuals did not perform above chance level in the condition meant to rule out that subjects relied on the heuristic: “avoid the sample drawn out of the population with more dis-preferred items”. In the children study, the methods of the experiments meant to rule out this quantity heuristic are ambiguous. In these experiments (Experiment 3

and 4), the authors made the assumption that a third type of object (green lollipops) inserted in the populations was considered as neutral by the children because they did not encounter it in the preference trial. If this was the case, results indicate that children did not try to avoid the sample drawn out of the population with more dis-preferred food. However, in case this assumption was not warranted, results would indicate the contrary. In the parrot study, no conditions were tested to rule out the use of quantity heuristics. Furthermore, as in all studies (except the parrot study) items were stacked on top of each other in the presented populations, subjects could have had the illusion that there were always more preferred items in the favourable population and could have relied on a comparison of absolute quantities of visible preferred items to solve the task. In no other study except ours were there any conditions that ruled out that individuals based their inferences on a comparison of visible preferred food items. Finally, as shown in a study with older children (Falk, Yudilevich-Assouline & Elstein, 2012), it is possible that individuals rely on different quantity heuristics interchangeably. This strategy would allow them to solve most tasks but does not require intuitions about probabilities. Looking at individual performance in the nonhuman primates studies, we see that there are only few subjects who were consistently good across all conditions, which could indicate that individuals used a mixture of different heuristics to solve the tasks.

It must also be noted that this paradigm in itself is not sufficient to provide answers about statistical reasoning abilities, as it only focused on inferences from populations to samples. However, there are other studies indicating that children are proficient in other aspects of statistical reasoning. For example, infants take into account the sampling process to form expectations about populations of objects and samples (Xu & Denison, 2009). They use samples to make inferences about sampling processes, which they in turn use to make generalisations about the properties of the samples (Gweon et al., 2010). They expect samples to be representative of populations (Xu & Garcia, 2008). These studies did, however, not consistently investigate whether individuals' expectations were shaped by the probabilities of different outcomes, or whether they relied on heuristics. Similarly, chimpanzees consider whether a sampling process is intentional or blind to form expectations about samples drawn out of populations (Eckert, Rakoczy, Call, Herrmann & Hanus, 2018). Apes also expect populations to be representative of samples (Eckert, Rakoczy & Call, 2017). However, it seems that these expectations are not based on an estimation of probabilities but rely on quantity heuristics.

Finally, these findings leave open some interesting questions regarding the usage of statistical reasoning abilities in everyday life scenarios. Reasoning about

random processes and making inferences from populations of objects to samples drawn out of them and vice versa is an ability that can help with making rational decisions in games of chance, for example, but no animals other than humans enjoy spending time in casinos or playing lotteries. Outside of games of chance, it is difficult to imagine a scenario involving random drawing and access to complete information about populations. In fact, in real life scenarios, information about populations is rarely obtainable. This is the reason why statistical reasoning is needed in the first place. Uncertainty usually stems from the need to rely on samples of data to learn about hidden properties of the environment. Is this tree producing good fruits? Is this individual cooperative? Is this pill creating headaches? One must rely on repeated interactions between organisms and their environment, to infer the state of these hidden properties. There are studies showing that infants, macaques, capuchin monkeys, and apes rely on statistical regularities of repeated types of events to make inferences about hidden properties of people or of objects (B. J. Edwards et al., 2014; Gweon & Schulz, 2011; Meyer, 1960; Völter, Sentís & Call, 2016; Waismeyer et al., 2015), however in these studies, absolute and relative frequencies of types of events are not disentangled, it is thus not clear whether subjects relied on heuristics or whether they estimated probabilities.

In summary, with the current evidence, we can conclude that some important aspects of statistical reasoning seem to be evolutionary ancient, such as considering the sampling process to make inferences from populations to samples and expecting samples to be representative of populations. There is evidence that children, apes, capuchin monkeys and a parrot rely on proportions of objects to make inferences from populations of objects to samples, but more research testing alternative hypotheses is needed. It is also not clear whether long-tailed macaques failed at making inferences from populations to samples because of a lack of probability intuitions or due to extraneous factors. Alternative methods that do not use food to convey numerical information nor humans as random processes should be developed, to reduce the demands of the tasks extraneous to statistical reasoning. This could in turn serve to increase the range of species that can be tested. What should also be investigated is how reasoning from populations of objects to samples and vice versa is extendable to reasoning about events.

5.1.2 Reasoning about repeated types of events

In my second study, I investigated whether long-tailed macaques rely on relative frequencies of repeated events to make rational decisions under uncertainty and I included different conditions to rule out that they rely on quantity heuristics. The

results of this study show a different picture of the statistical reasoning abilities of long-tailed macaques. In Experiment 1, it seems that long-tailed macaques monitored the different rewarding frequencies associated with different objects to predict which objects were leading to more probable rewards. We could rule out that monkeys based their choices on a mere comparison of absolute frequencies. Instead, our findings indicate that they compared relative frequencies, and therefore probabilities, to make predictions. They used these predictions to make rational decisions when several options were available simultaneously. Moreover, it seems that long-tailed macaques did not make predictions based solely on the relative frequencies of rewards associated with each option but also on the amount of information that was available for each option. Sample size therefore seems to matter to long-tailed macaques. Results of Experiment 2 and 3 further suggest that long-tailed macaques can monitor events of different levels of complexity and keep track of more complex covariation patterns to disentangle between factors that are relevant for their predictions and factors that are not, independent of whether these factors are social or physical. However, either the complexity of the event itself, or the nature of the different factors slowed down monkeys' learning process.

These findings raise some questions about statistical reasoning based on repeated types of events. The first question is whether what we call statistical reasoning here is actually only the result of associative learning with different schedules of reinforcement (Ferster & Skinner, 1957) or whether it truly results from a comparison of different probabilities of rewards. In Exp. 1a, 1b and 1c, monkeys' performance reached ceiling, which indicates that they made optimal choices based on probabilities, rather than engaged in probabilistic matching based on the different reinforcement schedules. Probabilistic matching was documented in many species, including fish (Behrend & Bitterman, 1961), pigeons (Bullock & Bitterman, 1962), rats (Hickson, 1961), rhesus monkeys (Wilson et al., 1964) and human children (Yost et al., 1962). However under different circumstances, rats, monkeys, children, and pigeons, also engage in optimal decisions (Meyer, 1960; Parducci & Polt, 1958; Roberts, MacDonald & Lo, 2018; Yost et al., 1962). It could be that rather than responding to reinforcement, individuals sometimes rely on win-stay, loose-switch strategies, which eventually leads to probabilistic matching, and sometimes they rely on probabilities. This seems to have been the case with long-tailed macaques in our experiments and was also discussed in Wilson et al. (1964) as a possible behaviour for rhesus macaques and rats. Nonetheless, in the future, to completely rule out associative learning processes, one should develop a set of tasks in which the reinforcement schedules associated with different options are the same,

but the probabilities of rewards differ.

A second question raised by the findings of my second study refers to the role mental representations and levels of abstractions play in statistical reasoning. For example, in Experiments 2 and 3, the probability associated with each option were the probability that each agent managed to open the box containing rewards. However, what one saw were two agents performing exactly the same behaviour, i.e., reaching towards the box, that resulted in different outcomes, i.e., the box opening or not. If one cannot represent agents as being more or less proficient at opening boxes, and boxes as being more or less easy to open, the task becomes much harder. In fact, in this case one has to remember all occasions on which agent A reached towards the box, and whether the box opened or not rather than remembering: “Agent A was successful 100% of the time, agent B only 40%”. In other words, rather than compressing one event into one datum such as: “the agent was successful”, one has to remember two data such as: “the agent reached towards the box; the box opened”. The same task can therefore be more or less demanding for different species, depending on the mental representations individuals possess rather than depending on their intuitions about probabilities. It is therefore possible that when we compare performance in statistical reasoning between species, we actually compare species with different levels of abstractions rather than species with different levels of intuitions about probabilities. It is also possible that mental representations affect the domains in which statistical reasoning can be applied.

In summary, the findings from my second study suggest that long-tailed macaques are statistical reasoners when provided with sequential events as information input. These findings also suggest that statistical reasoning might have been present in the last common ancestor of New World monkeys, Old world monkeys and apes after all, and might even be more evolutionary ancient, as parrots and pigeons (Roberts et al., 2018) recently tested in a similar paradigm as in my second study, also share this ability. The statistical reasoning ability of long-tailed macaques highlighted in my second study is also easier to extend to real life decision-making under uncertainty. Long-tailed macaques seem to possess a powerful cognitive mechanism to learn about their physical and social environment, that might only be restricted by the kind of abstract representations monkeys have.

5.1.3 Differences between different types of reasoning

In both my first and second study, long-tailed macaques had to choose between two options associated with different probabilities of reward. However, what is not clear is whether the cognitive mechanisms required to make rational decisions in the first

study are the same as the mechanisms required to make rational decisions in the second study. In a recent study by Roberts et al. (2018), pigeons were tested in tasks that were similar to Experiment 1 of my second study. Subjects could peck two different keys associated with different probabilities of reward, and from time to time, both keys were available simultaneously, so that subjects had to make a choice between both. Different conditions ruled out that pigeons relied on a comparison of absolute quantities of events rather than on relative frequencies. The authors in this study compare this ability to the abilities of children in Denison and Xu (2014), of apes in Rakoczy et al. (2014), and of capuchins in Tecwyn et al. (2017). Findings of my first and second study suggest that there might be different cognitive mechanisms at play, as the same group of long-tailed macaques did not perform at similar levels in both studies. However, as the performance of long-tailed macaques could have been hindered in the first study by factors extraneous to statistical reasoning, the difference in performance between my first two studies alone is not sufficient to infer that different mechanisms are at play.

Aside from differences in the kind of uncertainty that has to be dealt with in both kind of tasks (i.e., uncertainty inherent to random processes vs. uncertainty inherent to inductive inferences), there is another noteworthy difference: in the first study, subjects had to make inferences about single event probabilities, that is, probabilities that they had never experience before, whereas in the second study, subjects had to make inferences about experienced contingencies. In the first study, statistical information had to be extracted from the present state of the environment. In the second study, information had to be extracted from repeated types of events, inferences therefore relied on information from past events that individual had experienced. Research on statistical reasoning in human adults showed that subjects are better at making inferences when they experienced events compared to when they were provided with symbolic descriptions of problems (Bar-Hillel, 1980; Phillips & Edwards, 1966; Rehder & Waldmann, 2017). This suggests that experience can facilitate reasoning in some contexts (see Hertwig, Hogarth & Lejarraga, 2018, for an elaboration).

In Experiments 3 and 4 of Téglás et al. (2007), the relation between probability intuitions based on present information and on past frequencies was assessed. Three- and 5-year-olds saw short films of a ball randomly moving inside a rectangular box that had one opening on one side and three openings on the other side. After several seconds, the content of the box was occluded, and children were asked to press a button as soon as they saw the ball exit from one of the holes. Children underwent several sessions in which they experienced that the ball came more frequently out of

one opening rather than the three openings. Based on the present state of the environment, children should thus have had a prior belief that the ball would come out of one of the three openings (expected openings). Their reaction times of pressing the button were consistent with this hypothesis, as children were quicker when the ball came out of one of the expected openings than when it came out of the unexpected opening. However, across sessions, this prior belief should be updated based on the experienced frequencies. Reaction times of 5-year-olds changed across sessions, children became quicker when the ball came out of previously unexpected opening. This was not the case for 3-year-olds. They did not update their priors based on the experienced evidence. The authors concluded that: “far from being the foundation of probabilistic reasoning, experienced frequency moulds the expectation of future events only at a late stage of development” and “it looks like perceived frequencies and explicit reasoning about future states of affairs are computed by different mechanisms”.

It is surprising that three-year-olds do not learn from experienced frequencies, whereas 1-year-old long-tailed macaques and pigeons do. But actually, what Téglás et al. (2007) compared in their study were not probability estimations based on present information to estimations based on experienced frequencies. What they assessed was whether children updated their priors based on new evidence. It is possible that 3-year-olds do learn from experienced frequencies when their priors are flat, as do long-tailed macaques, but that they have difficulties updating their priors if they are not flat. These findings could therefore also indicate that both kinds of intuitions are present in children from birth and are evolutionary ancient, but that the ability to update prior beliefs based on new information might emerge later during development. Nonetheless, results from Téglás et al. (2007) suggest that there are different cognitive mechanisms that are involved in assessing probabilities of uncertain events.

In summary, findings from my first and second study, as well as recent findings with apes, capuchins, parrots, and pigeons, suggest that statistical reasoning is evolutionary ancient. Different cognitive mechanisms could be at the origins of probability intuitions and they might have followed different evolutionary trajectories. The ability to make inductive inferences from samples of experienced events might have already been present in the last common ancestor between birds and mammals, some 300 million years ago (Burt et al., 1999), or might have evolved in convergent fashion in birds and primates. However, the ability to make predictions from populations of objects to samples might only have evolved in New World monkeys, Apes and birds, but more studies are needed to confirm this.

5.2 Of smart babies and dumb preschoolers

In my third study I investigated whether the performance of infants and preschoolers at making inferences from populations of objects to samples by relying on proportions of objects had been respectively over-estimated and under-estimated in previous studies. Results of my third study showed that preschoolers do not rely on proportions of objects to make predictions from populations to samples, as was already suggested by previous research (Giroto et al., 2016). Our results with infants are harder to interpret, as subjects also failed in the baseline condition, in which outcomes were certain. There are two explanations for this: either the procedure confused them or they did not have a strong preference for one of the objects. It is unclear why they would be confused by the procedure, as we tried to keep it as close as possible to the procedure used in Denison and Xu (2014) in which infants were successful. It is also unclear why they would have a preference for one lollipop over another due to different colours, but not for a blue egg with cartoon eyes containing something over a white egg without eyes and content. Another possibility is that the performance of infants in Denison and Xu (2014) has been over-estimated and that subjects, instead of estimating probabilities, based their choices on slight indications from their parents who were not blind to the conditions. This last possibility is in line with the findings of Téglás et al. (2015), namely that infants are able to estimate probabilities of events in an extensional way but not by relying on proportions of large quantities of objects.

There are other findings that suggest that infants rely on proportions of objects to estimate probabilities of random drawing. For example, in Xu and Garcia (2008) infants expected samples drawn out of hidden populations of objects to be representative of the populations in terms of their proportions of red and white objects. In Gweon et al. (2010) infants relied on the probability of obtaining a given sample drawn out of a population of objects to assess whether the drawing was intentional or not. However, in these studies nothing ruled out that children relied on quantity heuristics to make predictions.

Infants seem to be intuitive statisticians insofar as they consider cues of intentional or random drawing to make predictions, as they expect samples to be representative of populations, and as they make inferences about sampling process based on evidence. Whether they rely on proportions of objects to assess probabilities of random drawing is yet not clear. It is possible that infants' performance has been overestimated and that they rely on a series of quantity heuristics to make predictions about random drawing when they cannot estimate probabilities in an extensional way. With age, however, children develop their understanding of ran-

dom drawing, random distributions and combinatorics, and rely less and less on heuristics, as suggested by the work of Falk et al. (2012) and Piaget and Inhelder (1975). More studies are needed in which the preference of children for different object types is assessed with more care, and in which parents are blind to the experiments to shed more light on the developmental trajectory of probability intuitions. What also remains unclear is how this relates to children's abilities to reason about repeated types of events. Findings of Gweon and Schulz (2011) and Waismeyer et al. (2015) suggest that infants consider covariation patterns and relative frequencies of types of events to make predictions. However, these studies did not systematically rule out that children relied on quantity heuristics rather than probabilities.

In summary, findings from many different studies suggest that human infants are statistical reasoners, as they seem able to make inferences from populations to samples and from samples to populations taking sampling processes and sample sizes into account. However, more studies are necessary to assess whether children rely on an estimation of different probabilities of outcomes, or whether they rely on quantity heuristics to make predictions about random drawing. It could be that this ability only emerges later during childhood.

5.3 Future perspectives

As in any other field of research, there is room for improvement in research about statistical reasoning. More species should be studied to obtain a clearer picture of the evolutionary origins of this ability. Studies should include conditions that rule out the use of all possible heuristics. This has not been done carefully enough in the past. The demands of the tasks extraneous to statistical reasoning should be reduced to a minimum to avoid their interference with subjects' performance. Most importantly, in my opinion, more thoughts should be given to the nature of statistical reasoning. Are we talking of one or of several mechanisms? What is the difference between the different terms used to refer to the ability to make predictions under uncertainty, such as probabilistic reasoning, intuitive statistic, statistical reasoning, Bayesian reasoning? What are the requirements to call an individual, or a group of individuals, statistical reasoners? What role do mental representations play in statistical reasoning?

5.3.1 Definitions

A good start to answer some of these questions would be to provide more precise definitions based on probability and inferential statistics theory. As mentioned in

the introduction of this dissertation, probability theory provides mathematical tools to compute probabilities of outcomes of random processes. The uncertainty that is dealt with in probabilistic experiments is inherent to random processes, as no previous information indicates what is going to happen. Outcomes of random processes are estimated based on the sample space and on the probability associated with each type of outcomes.

Inferential statistics deals with uncertainty that is inherent to inductive inferences, that is, to generalisations made from samples of data. Randomness here comes into play at the moment of sampling. As we saw in the introduction, to increase the probability that a sample is representative of the population it stems from, it should be sampled randomly or without any bias.

One could therefore say that probabilistic reasoning refers to situations in which one has to make predictions about the outcomes of a random process and has perfect knowledge regarding all the possible outcomes and their respective probability. Or in other words, one has knowledge about a population and makes predictions about samples.

In contrast, one could say that statistical reasoning refers to situations in which information about populations is not available and in which the only accessible data comes in the form of samples drawn out of the population. Statistical reasoners should rely on statistical regularities observed in these samples to make inferences about populations. They should incorporate the size of the observed samples into their inferences, as well as the sampling process. They should adjust their inferences when provided with new data. In other words a statistical reasoner has knowledge about samples and makes inferences about populations.

Consistent with these definitions, Prasanta S. Bandyopadhyay and Steve Cherry (2011) wrote in their article entitled: “Elementary Probability and Statistics: A Primer”:

In short, drawing an inference from a sample to a population is statistics and drawing an inference from a population to a sample is mathematics that rests on using the theory of probability. (p.62)

One could also say that the result of statistical reasoning is learning and the result of probabilistic reasoning is a prediction. Often, both types of reasoning work together, as one first learns about the probability distribution of a given parameter, and then uses this knowledge to make new predictions about events in order to direct decisions. Both statistical and probabilistic reasoning therefore help making predictions and rational decisions under uncertainty inherent to inductive generalisations, but only probabilistic reasoning helps making predictions about outcomes

of random processes.

Is it ever the case, in real life, that one has perfect knowledge about the sample space of a future event and the probability associated with each possibility or are real life problems only problems of inferential statistics? It is difficult to provide a direct answer to this question as one should find examples of real life scenarios that are treated like probabilistic experiments. However, research on intuitions about probability can provide an indirect answer. For example, is it not the case in Téglás et al. (2007) that infants made probabilistic inferences based on the possible outcomes of a random process? If this had been a statistical problem, infants should first have learned the probability associated with each option through experience to be able to make rational predictions. As shown in Experiments 3 and 4 of Téglás et al. (2007), 5-year-olds did also integrate experienced frequencies into their predictions, but not 3-year-olds. These findings suggest that infants make inferences about random processes without any prior knowledge, based on the probability of each possibility.

Similarly, in Denison and Xu (2014), Rakoczy et al. (2014), Tecwyn et al. (2017), Clements et al. (2018), it seems that individuals made inferences from populations to samples based on an assumption of random drawing. If this had been a statistical problem, individuals should have updated their priors about the random sampling based on the outcomes. In fact, individuals in the nonhuman primates studies should have learnt that the sampling was not random, as the experimenter always drew an item of the majority type out of each population, but no learning was observed in any experiment. If we assume that individuals relied on proportions of objects and on an indication of random drawing to make inferences, we can say that these predictions were the result of probabilistic reasoning but not of statistical reasoning. Whether there are real life scenarios similar to probabilistic experiments is hard to answer but it seems that some primates brains have the ability to deal with probabilistic experiments.

Experiments in Xu and Garcia (2008), in contrast, are examples of statistical reasoning, if we assume that children relied on proportions of objects to make inferences about the content of the populations. Findings from Xu and Denison (2009), Eckert, Rakoczy et al. (2018) and from my second study would be examples of statistical reasoning and probabilistic reasoning, assuming again that individuals relied on relative frequencies of objects in the children study, as subjects first learnt something, either about the sampling process in the two first studies or about the probability of rewards associated to each option in my study, and then used this knowledge to make new predictions and rational decisions.

What are intuitive statistics and Bayesian reasoning? Intuitive statistics should be a term used to refer to statistical reasoning abilities that have not been acquired through education, which is the case of studies with young human children, non-human primates and prenumerate human adults. Bayesian reasoning should more specifically refer to the ability to update prior beliefs based on evidence following Bayes' theorem, as was illustrated in the example of Phillips and Edwards (1966) mentioned in the introduction. The inferences made by infants in Xu and Garcia (2008) could be seen as Bayesian reasoning, as infants were first shown two possibilities of populations of objects, and then, based on samples drawn out of a hidden population, they formed a belief about the possible state of the hidden population. However, methods of this study did not allow to assess whether infants distinguished between the impact of the evidence (i.e., which hypothesis was best supported by the evidence), and the posterior probability of each hypothesis, as the hypothesis that was best supported by the evidence was also the hypothesis with the highest posterior probability (this is always the case when priors are flat). Remember that in the example of the Green and the Blue cab companies (Bar-Hillel, 1980), it was ruled out that adults were statistical reasoners because they did not include the base rate at which each type of car occurred in the city into their calculations. Findings of Xu and Garcia (2008) are not sufficient to conclude that infants did. There is to my knowledge no research with infants and nonhuman primates that clearly showed that infants and nonhuman primates update priors based on evidence following Bayes' rule.

As we saw in the previous section, different mechanisms might be responsible for different probability intuitions. To shed more light on the nature and the evolution of probability intuitions, it therefore seems warranted to name different abilities differently and to compare performance of animals with tasks that have the same computational difficulty and that provided the same kind of information.

5.3.2 Considering mental representations

As I already mentioned above, it is hard to tell apart the role played by mental representations and the role played by probability intuitions in the performance of subjects in statistical reasoning tasks. One of the prerequisite to engage in statistical reasoning is the ability to form hypotheses. This is well formulated in a paper from Tenenbaum et al. (2011):

If the mind goes beyond the data given, another source of information must make up the difference. Some more abstract background know-

ledge must generate and delimit the hypotheses learners consider, or meaningful generalization would be impossible. (p. 1279)

If a reasoner is unable to represent a given probability distribution, a relevant factor or variable, no hypothesis can be formed and no learning will ensue, irrespective of whether the reasoner is a statistical reasoner or not. One can represent a factor or a variable more or less abstractedly, and this in turn can affect one's ability to generalise. Take the example of our two agents in Experiment 3 of my second study, who opened different boxes. As was already mentioned above, representing the probability distribution of the successes and failures to open the box as different levels of proficiency can help in memorizing the events over time. Now imagine that new boxes with different colours and shapes were placed in front of the two humans, and monkeys had to choose between the two humans, as previously, but without having seen the humans interact with the new boxes. If long-tailed macaques can represent the different boxes as being part of a bigger category, such as "objects that can be opened in similar ways", and if they can represent the previous successes and failures of the agents as a constant characteristics of these agents, then they should be able to generalise what they learnt in the previous trials to a new context. If they cannot represent these variables, they would have to go through a new sampling phase to learn the probabilistic contingencies between human, box and reward in this new context.

Another way how higher levels of abstractions can enhance statistical reasoning is by making reasoners more aware of conditional factors (Cheng, 1993). Take again the example of my first study and similar paradigms. If a reasoner cannot represent a blind agent sampling items from a box as a random sampling process, she will not use it as a condition for considering all items in the boxes as having equal probability of being drawn and will fail in this task, independent of whether she has intuitions about probabilities or not. The ability to represent some factors as potentially predictive of events and not others can also limit the number of computations the brain needs to do in order to find meaningful regularities (Tenenbaum et al., 2011). Imagine that long-tailed macaques in our experiments also considered the table on which the objects were placed, the side on which they were placed, the order in which objects were presented, etc. as potential factors affecting the probability to obtain a reward. In some trials, some of these factors would covary with the higher frequency of rewards, monkeys would therefore need more trials and more time to rule these factors out as potential predictors. A task that is easy for some animals can be hard for others.

Statistical reasoning might be content-free in the sense that it uses probability

distributions as input information, and probability distributions can be derived from any repetition of types of events and from any proportion of objects over space and time. However, whether brains can derive these probability distributions in the first place might depend on the available mental representations. To shed more light on the evolution of statistical reasoning, it seems therefore important to investigate how different mental representations can influence inferences. What sets humans apart from other species might be their ability to compute more complex probabilities, but it might also be the presence in their brains of more abstract representations of the world. Both factors might be at the origin of human's larger-scale systems of knowledge (Tenenbaum et al., 2011).

5.4 Conclusion

In summary, the findings from my first two studies, in conjunction with findings from studies on other animals, suggest that intuitions about probabilities are older than the Hominidae family. Different species of primates and birds seem to share similar probability intuitions. However, different cognitive mechanisms might be at the origins of intuitions about probabilities. One mechanism might be responsible for making inferences about never-experienced events and to use statistical information extracted from the present state of the environment as input information. The other mechanism might rely on experienced frequencies and to use statistical information extracted from past events. The fact that long-tailed macaques failed at making inferences from populations to samples but made rational decisions when they could extract statistical information from repeated types of events suggests that reasoning about experienced events might be evolutionary older than reasoning about never-experienced outcomes. The findings from my third study confirm that preschoolers do not rely on proportions to make predictions from populations to samples, and further suggest that infants' abilities might have been overestimated and that proportional reasoning only develops around 5 years of age. However, this conclusion has to be treated carefully, as other factors might have impaired infants' performance. In the future, it would be interesting to classify the different tasks used to test probability intuitions into different categories, depending on their computational difficulty, on the mental representation a reasoner needs to solve these tasks, and the type of information provided. Such a classification could foster better comparisons between performances of animals or of different ages groups and better inform the evolutionary origins and developmental trajectory of statistical reasoning.

References

- Bandyopadhyay, P. S. & Cherry, S. (2011). Elementary probability and statistics: A primer. In P. S. Bandyopadhyay & M. Forster (Eds.), *Philosophy of statistics* (Vol. 7, pp. 53–95). Oxford, UK: Elsevier.
- Bandyopadhyay, P. S. & Forster, M. R. (2011). Philosophy of statistics: an introduction. In P. S. Bandyopadhyay & M. Forster (Eds.), *Philosophy of statistics* (Vol. 7, pp. 1–50). Oxford, UK: Elsevier.
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*, 211–233.
- Batanero, C., Green, D. R. & Serrano, L. R. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematical Education in Science and Technology*, *29*, 113–123.
- Behrend, E. R. & Bitterman, M. (1961). Probability-matching in the fish. *The American Journal of Psychology*, *74*, 542–551.
- Biernaskie, J. M., Walker, S. C. & Gegeer, R. J. (2009). Bumblebees learn to forage like bayesians. *The American Naturalist*, *174*, 413–423.
- Bitterman, M., Wodinsky, J. & Candland, D. K. (1958). Some comparative psychology. *The American Journal of Psychology*, *71*, 94–110.
- 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*, 82–86.
- Braine, M. D. (1962). Piaget on reasoning: A methodological critique and alternative proposals. *Monographs of the Society for Research in Child Development*, *27*, 41–63.
- Brainerd, C. J. (1981). Working memory and the developmental analysis of probability judgment. *Psychological Review*, *88*, 463–502.
- Bryant, P. & Nunes, T. (2012). *Children's understanding of probability: A literature review (summary report)*. London, UK: Nuffield Foundation.

- Bullock, D. H. & Bitterman, M. (1962). Probability-matching in the pigeon. *The American Journal of Psychology*, *75*, 634–639.
- Burt, D. W., Bruley, C., Dunn, I. C., Jones, C. T., Ramage, A., Law, A. S., ... others (1999). The dynamics of chromosome evolution in birds and mammals. *Nature*, *402*, 411–413.
- Calude, C. S., Dinneen, M. J., Dumitrescu, M. & Svozil, K. (2010). Experimental evidence of quantum randomness incomputability. *Physical Review A*, *82*, 022102.
- Carlson, S. M., Davis, A. C. & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science*, *16*, 609–616.
- Cheng, P. W. (1993). Separating causal laws from casual facts: Pressing the limits of statistical relevance. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 215–264). Cambridge, MA: Academic Press.
- Cheng, P. W. & Lu, H. (2017). Causal invariance as an essential constraint for creating a causal representation of the world: Generalizing. In M. R. Waldmann (Ed.), *The oxford handbook of causal reasoning* (pp. 65–84). Oxford, UK: Oxford University Press.
- 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*, 166–177.
- Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1–73.
- Daston, L. (1988). *Classical probability in the enlightenment*. Princeton, NJ: Princeton University Press.
- Debnath, L. & Basu, K. (2015). A short history of probability theory and its applications. *International Journal of Mathematical Education in Science and Technology*, *46*, 13–39.
- 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*, 243–249.
- Denison, S. & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, *130*, 335–347.
- 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.

- 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*, e22693.
- Eckert, J., Rakoczy, H., Call, J., Herrmann, E. & Hanus, D. (2018). Chimpanzees consider humans' psychological states when drawing statistical inferences. *Current Biology*, *28*, 1959–1963.
- Edwards, B. J., Rottman, B. M., Shankar, M., Betzler, R., Chituc, V., Rodriguez, R., ... Santos, L. R. (2014). Do capuchin monkeys (*Cebus apella*) diagnose causal relations in the absence of a direct reward? *PloS One*, *9*, e88595.
- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: New York: Wiley.
- Falk, R., Yudilevich-Assouline, P. & Elstein, A. (2012). Children's concept of probability as inferred from their binary choices—revisited. *Educational Studies in Mathematics*, *81*, 207–233.
- Feigenson, L. & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, *91*, 173–190.
- Ferster, C. B. & Skinner, B. F. (1957). *Schedules of reinforcement*. East Norwalk, CT: Appleton-Century-Crofts.
- Fontanari, L., Gonzalez, M., Vallortigara, G. & Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. *Proceedings of the National Academy of Sciences*, *111*, 17075–17080.
- Gigerenzer, G. (1996). The psychology of good judgment: frequency formats and simple algorithms. *Medical Decision Making*, *16*, 273–280.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: frequency formats. *Psychological Review*, *102*, 684–704.
- Gigerenzer, G. & Hoffrage, U. (1999). Overcoming difficulties in bayesian reasoning: A reply to lewis and keren (1999) and mellers and mcgraw (1999). *Psychological Review*, *106*, 684–704.
- 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.
- Girotto, V. & Gonzalez, M. (2008). Children's understanding of posterior probability. *Cognition*, *106*, 325–344.
- Gweon, H. & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science*, *332*, 1524–1524.
- Gweon, H., Tenenbaum, J. B. & Schulz, L. E. (2010). Infants consider both the

- sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, *107*, 9066–9071.
- Hertwig, R., Hogarth, R. M. & Lejarraga, T. (2018). Experience and description: Exploring two paths to knowledge. *Current Directions in Psychological Science*, *27*, 123–128.
- Hickson, R. H. (1961). Response probability in a two-choice learning situation with varying probability of reinforcement. *Journal of Experimental Psychology*, *62*, 138–144.
- Hájek, A. (2012). Interpretations of probability. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2012 ed.). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2012/entries/probability-interpret/>.
- Johnson-Laird, P., Legrenzi, P., Girotto, V., Legrenzi, M. S. & Caverni, J.-P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological Review*, *106*, 62.
- Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.
- Kolmogorov, A. N. (1933). *Foundations of the theory of probability* (N. Morrison, Trans.). Oxford, UK: Chelsea Publishing Company.
- Kummer, H., Anzenberger, G. & Hemelrijk, C. K. (1996). Hiding and perspective taking in long-tailed macaques (*macaca fascicularis*). *Journal of Comparative Psychology*, *110*, 97–102.
- Laplace, P. S. (1812). *Théorie analytique des probabilités*. Paris, France: Courcier.
- Lima, S. L. (1984). Downy woodpecker foraging behavior: efficient sampling in simple stochastic environments. *Ecology*, *65*, 166–174.
- Lima, S. L. (1985). Sampling behavior of starlings foraging in simple patchy environments. *Behavioral Ecology and Sociobiology*, *16*, 135–142.
- Lindley, D. V. (2000). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*, 293–337.
- Longo, N. (1964). Probability-learning and habit-reversal in the cockroach. *The American Journal of Psychology*, *77*, 29–41.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: W. H. Freeman and Company.
- McNamara, J. M., Green, R. F. & Olsson, O. (2006). Bayes' theorem and its

- applications in animal behaviour. *Oikos*, *112*(2), 243–251.
- Meyer, D. R. (1960). The effects of differential probabilities of reinforcement on discrimination learning by monkeys. *Journal of Comparative and Physiological Psychology*, *53*, 173–175.
- Munakata, Y., McClelland, J. L., Johnson, M. H. & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*, *104*, 686–713.
- Parducci, A. & Polt, J. (1958). Correction vs. noncorrection with changing reinforcement schedules. *Journal of Comparative and Physiological Psychology*, *51*, 492–495.
- Pasquini, E. S., Corriveau, K. H., Koenig, M. & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*, 1216–1226.
- Phillips, L. D. & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*, 346–354.
- Piaget, J. & Inhelder, B. (1975). *The origin of the idea of chance in children.*(trans l. leake et al). Oxford, UK: WW Norton.
- Pighin, S., Girotto, V. & Tentori, K. (2017). Children’s quantitative bayesian inferences from natural frequencies and number of chances. *Cognition*, *168*, 164–175.
- Placì, S., Eckert, J., Rakoczy, H. & Fischer, J. (2018). Long-tailed macaques (*macaca fascicularis*) can use simple heuristics but fail at drawing statistical inferences from populations to samples. *Royal Society Open Science*, *5*, 181025.
- Rakoczy, H., Clüver, A., Saucke, L., Stoffregen, N., Gräbener, A., Migura, J. & Call, J. (2014). Apes are intuitive statisticians. *Cognition*, *131*, 60–68.
- Real, L. A. (1991). Animal choice behavior and the evolution of cognitive architecture. *Science*, *253*, 980–986.
- Rehder, B. & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & cognition*, *45*, 245–260.
- Roberts, W. A., MacDonald, H. & Lo, K. H. (2018). Pigeons play the percentages: computation of probability in a bird. *Animal Cognition*, *21*, 575–581.
- Savage, L. J. (1972). *The foundations of statistics*. New York, NY: Dover Publications, Inc.
- Schmitt, V. & Fischer, J. (2011). Representational format determines numerical competence in monkeys. *Nature Communications*, *2*, 257.
- Sobel, D. M., Tenenbaum, J. B. & Gopnik, A. (2004). Children’s causal infer-

- ences from indirect evidence: Backwards blocking and bayesian reasoning in preschoolers. *Cognitive Science*, *28*, 303–333.
- Tecwyn, E. C., Denison, S., Messer, E. J. & Buchsbaum, D. (2017). Intuitive probabilistic inference in capuchin monkeys. *Animal Cognition*, *20*, 243–256.
- Téglás, E., Girotto, V., Gonzalez, M. & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences*, *104*, 19156–19159.
- 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.
- 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*, 1054–1059.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285.
- Valone, T. J. (2006). Are animals capable of bayesian updating? an empirical review. *Oikos*, *112*, 252–259.
- Varberg, D. E. (1963). The development of modern statistics. *The Mathematics Teacher*, *56*, 252–257.
- Völter, C. J., Sentís, I. & Call, J. (2016). Great apes and children infer causal relations from patterns of variation and covariation. *Cognition*, *155*, 30–43.
- Waismeyer, A., Meltzoff, A. N. & Gopnik, A. (2015). Causal learning from probabilistic events in 24-month-olds: an action measure. *Developmental Science*, *18*, 175–182.
- Wilson, W. A., Oscar Jr, M. & Bitterman, M. (1964). Probability-learning in the monkey. *Quarterly Journal of Experimental Psychology*, *16*, 163–165.
- Xu, F. & Arriaga, R. I. (2007). Number discrimination in 10-month-old infants. *British Journal of developmental psychology*, *25*, 103–108.
- 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*, *105*, 5012–5015.
- Xu, F. & Kushnir, T. (2013). Infants are rational constructivist learners. *Current Directions in Psychological Science*, *22*, 28–32.
- Xu, F., Kushnir, T. & Benson, J. B. (Eds.). (2012). *Rational constructivism in cognitive development* (Vol. 43). Cambridge, MA: Academic Press.
- Yost, P. A., Siegel, A. E. & Andrews, J. M. (1962). Nonverbal probability judgments

by young children. *Child Development*, 33, 769–780.

Appendix **A**

Supplementary material for: Do children and preschoolers quantify probabilities based on proportions?

Do infants and preschoolers quantify probabilities based on proportions?

Sarah Placi^{1,2,3*}, Julia Fischer^{1,3*}, Hannes Rakoczy^{2,3*}

¹Cognitive Ethology Laboratory, German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany

²Department of Developmental Psychology, University of Göttingen, Waldweg 26, 37073 Göttingen, Germany

³Leibniz ScienceCampus Primate Cognition, German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany

*Corresponding Author: Sarah Placi

*Equal contribution

Email: sarah.placi@gmail.com

Phone +49 551 3851-0

Keywords: Probabilistic reasoning; Developmental psychology; Intuitive statistics; Child cognition

Supplementary material

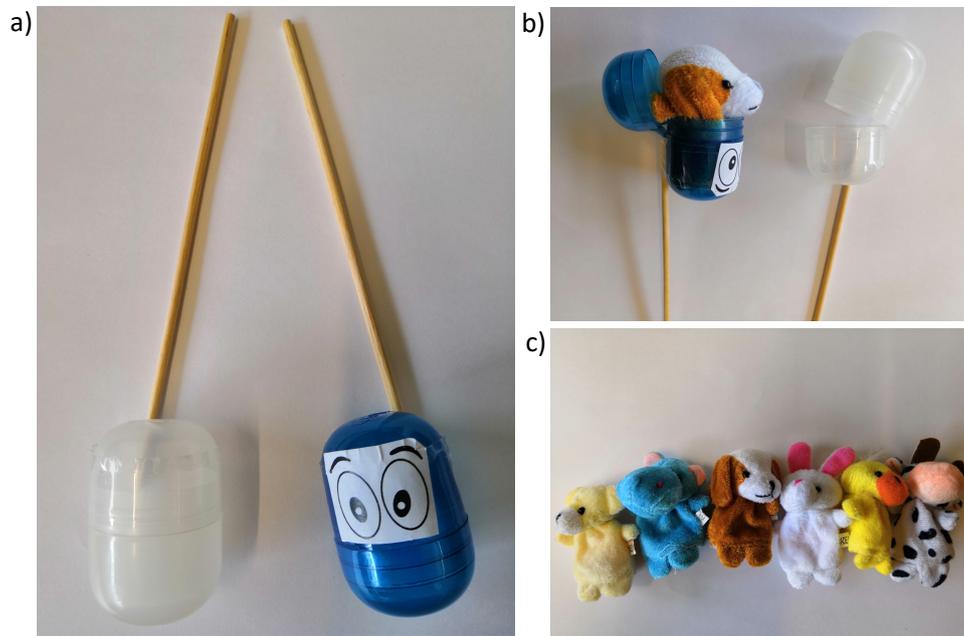


Figure S1. Objects used as rewards during the study. a) Children were presented with white and blue Kinder egg capsules. On the blue eggs were also displayed cartoon eyes. b) All blue eggs contained a fingerpuppet, all white eggs were empty. c) Different fingerpuppets used during the study.