Cognition 131 (2014) 60–68

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Apes are intuitive statisticians

Hannes Rakoczy [a,*], Annette Clüver [a], Liane Saucke [a], Nicole Stoffregen [a], Alice Gräbener [a], Judith Migura [a], Josep Call [b]

[a] *Institute of Psychology & Courant Research Centre "Evolution of Social Behaviour", University of Göttingen, Germany*
[b] *Max-Planck-Institute for Evolutionary Anthropology, Department of Developmental and Comparative Psychology, Leipzig, Germany*

ARTICLE INFO

Article history:
Received 13 March 2013
Revised 12 December 2013
Accepted 19 December 2013

Keywords:
Comparative psychology
Primate cognition
Intuitive statistics
Numerical cognition

ABSTRACT

Inductive learning and reasoning, as we use it both in everyday life and in science, is characterized by flexible inferences based on statistical information: inferences from populations to samples and vice versa. Many forms of such statistical reasoning have been found to develop late in human ontogeny, depending on formal education and language, and to be fragile even in adults. New revolutionary research, however, suggests that even preverbal human infants make use of intuitive statistics. Here, we conducted the first investigation of such intuitive statistical reasoning with non-human primates. In a series of 7 experiments, Bonobos, Chimpanzees, Gorillas and Orangutans drew flexible statistical inferences from populations to samples. These inferences, furthermore, were truly based on statistical information regarding the relative frequency distributions in a population, and not on absolute frequencies. Intuitive statistics in its most basic form is thus an evolutionarily more ancient rather than a uniquely human capacity.

1. Introduction

Much research has suggested that reasoning about probabilities develops late in ontogeny, depends on language and formal education (Piaget & Inhelder, 1975), remains fragile even in adulthood (Tversky & Kahneman, 1974, 1981), and only works under special circumstances (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). Exciting new research, however, suggests that such reasoning capacities might well be in place in the absence of language. Even preverbal infants engage in some intuitive statistics: they expect randomly drawn samples to reflect the distribution in the population drawn from and vice versa (Denison & Xu, 2010b; Téglás, Girotto, Gonzalez, & Bonatti, 2007; Xu & Garcia, 2008).

This early intuitive statistics seems to be a cognitive capacity that is functionally integrated in humans with other cognitive domains from very early on: For example, infants already integrate information about physical and psychological background conditions into their statistical inferences when judging whether sampling processes are random or non-random. Regarding physical information, for example, infants understand that mechanical constraints (e.g. some kinds of objects in a population cannot be drawn physically in the same way as others) can turn a sampling process into a non-random one such that the sample need not reflect the distribution in the population (Denison & Xu, 2010a; Téglás et al., 2007). Moreover, statistical information is combined with geometrical and temporal information in rather systematic ways to form predictions about future events (Téglás et al., 2011). Regarding psychological information, infants appreciate that when a person draws from a population but has both a preference regarding the different kinds of objects in the population and visual access, her sampling will probably be non-random and her sample will thus not match the distribution of the population (Xu & Garcia, 2008).

---

* Corresponding author. Address: Department of Developmental Psychology & Courant Research Centre "Evolution of Social Behaviour", University of Göttingen, Waldweg 26, D-37073 Göttingen, Germany. Tel.: +49 551 399235.

E-mail address: hannes.rakoczy@psych.uni-goettingen.de (H. Rakoczy).

0010-0277/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.12.011

Conversely, infants draw inferences in the other direction, from statistical to psychological states of affairs: when confronted with a person who draws samples that are absolutely non-representative of the populations, infants assume the person must have informational access and corresponding preferences (Kushnir, Xu, & Wellman, 2010; Ma & Xu, 2011). Convergent findings have shown such intuitively statistical expectations in infants and toddlers with a number of different measures tapping different types of behaviors: looking time in response to violations of expectations (e.g. Xu & Garcia, 2008), active choice measures (of samples drawn from different populations; (Denison & Xu, 2010b) and actions directed towards others (such as giving them the kind of item they prefer; (Kushnir et al., 2010; Ma & Xu, 2011).

Compared to the information available for human infants, nothing is currently known about the phylogenetic origins and distributions of such intuitive statistics. We do not know how old evolutionarily capacities for intuitive statistics are, and we do not know whether they are shared by any non-human animals. Studies on optimal choice and foraging (Balci, Freestone, & Gallistel, 2009; Kamil, Krebs, & Pulliam, 1987; Stephens, 2008; Stüttgen, Yildiz, & Güntürkün, 2011) and numerosity discrimination (Brannon & Terrace, 1998; Hanus & Call, 2007) have demonstrated that non-human animals share with humans basic cognitive capacities to maximize the amount of food rewards on the basis of perceptual information. Those studies, however, lack some of the crucial features present in intuitive statistics research. Unlike subjects in optimal choice studies, subjects in intuitive statistics studies form expectations and select optimally based on statistical information without any prior training to associate the stimuli and their reinforcement contingencies or any other reliance on past sampling (Téglás et al., 2007, 2011).

It is true that such good first trial performance in the absence of training can also be found in primates' discrimination of absolute set sizes. First, from numerous comparative studies we know that many non-human animals, notably primates, share with humans an analog magnitude system that allows for the approximate discrimination between arbitrarily large sets (Cantlon & Brannon, 2006, 2007; Flombaum, Junge, & Hauser, 2005). The signature limit of this capacity, following Weber's Law, is constituted by the ratios of the sizes of two sets to be discriminated: if a subject can discriminate 4 from 8 objects, it can discriminate 10 from 20, 150 from 300, etc. Second, humans and other primates share an object individuation system that allows for the exact parallel individuation ("subitizing") of small sets (Hauser, MacNeilage, & Ware, 1996). The signature limit here is defined by the absolute set sizes: only sets smaller than 3 (infants) or 4 (monkeys and apes) can be discriminated, such as 1:2, 2:3, and 1:3 (see (Carey, 2009), for review). In contrast to such tasks, however, intuitive statistical problems crucially require representing truly statistical matters, namely relative rather than absolute frequencies – that is, frequencies of items of a given kind in a population (say, winner tickets in a lottery) relative to the frequencies of all kinds of items in the population (all tickets). It is thus an open question whether intuitive statistical reasoning, understood as the capacity to flexibly draw inferences from populations to samples and vice versa, is evolutionarily recent and uniquely human or evolutionarily ancient and shared with other animals. Here we report a series of studies that speaks to that question. These studies with our closest relatives, the great apes, investigated one of the most basic forms of such intuitive statistical capacities: the ability to draw inferences from information about a population to a randomly drawn sample. We used tasks modeled after those developed in recent infant studies (Denison & Xu, 2010b). In these tasks, subjects are confronted with two visible populations with different distributions of items of two kinds (one preferable over the other) and the experimenter randomly draws from each population a 1-object-sample that the subject cannot see. Subjects are then given a choice between the two samples. These tasks thus require the subjects, first, to distinguish between the two populations according to the ratios of the two kinds of objects in their distributions and, second, to form expectations about the probability of sampling events accordingly, that is, expectations as to which sample is more likely to contain an object of the more desirable kind.

Control experiments ruled out alternative explanations such as simpler choice heuristics (Exp. 2 and 3), Clever Hans effects (Exp. 5 and 6) and use of olfactory information (Exp. 7). Most importantly, two experiments (Exp. 4 and 6) tested whether such inferences were truly based on probability information and not just on information about absolute frequencies.

2. Experiment 1: inferences from populations to samples

2.1. Subjects

Participants in all experiments were recruited from a group of four species of Great Apes ($N = 33$; Female $N = 24$): Chimpanzees (*Pan troglodytes verus*, $N = 17$), Gorillas (*Gorilla gorilla*, $N = 5$), Orangutans (*Pongo pygmaeus*, $N = 6$) and Bonobos (*Pan paniscus*, $N = 5$) housed at the Wolfgang Köhler Primate Research Center (WKPRC) in the Leipzig Zoo. Mean age of animals was 16;10 (years; months) with a range of 6;2-30;6. About one third were hand-reared and the remaining two third**s** were mother-reared. All subjects had experience in cognitive studies and were used to receiving food-items as reinforcement (see SI Table 1 for a detailed description of the animals' demographics and background). 28 apes (15 Chimpanzees, 2 Gorillas, 6 Orangutans and 5 Bonobos) were included in the final sample of this experiment. Four further apes (2 Chimpanzees and 2 Gorillas) were tested but excluded from data analysis due to inconsistent item preference during the Preference Test ($N = 1$) or because they did not complete all trials due to lack of motivation ($N = 3$).

2.2. Design and procedure

Populations of banana pellets and carrot pieces were presented in two transparent buckets. Both buckets contained the same absolute amount of food items (80), with distribution of banana pellets to carrots of 4:1 in bucket A and 1:4 in bucket B. Each ape participated in 12 trials in which an experimenter drew one item from each bucket (always of the majority type). Apes were tested individually by two experimenters in special testing cages or their sleeping quarters. Stimuli were presented on a table (35 × 78 cm) mounted to one side of the testing cages. A Plexiglas panel mounted on the cage mesh and perpendicular to the table separated the ape from the experimenters. Two small holes (∅2 cm; distance between holes 59 cm) drilled into the Plexiglas panel allowed the apes to indicate their choices by inserting a finger into one of them.

2.2.1. Preference test

For each experiment, a preference test was administered place before the first test trial session. One banana pellet and one carrot piece were placed in front of the holes on the experimenters' side of the panel. Apes indicated their choice with their finger and immediately received the selected food item as reinforcement. This preference test was carried out twice in succession to establish whether apes' preference was consistent. Virtually all apes (with the exception of one subject in Exp. 1, and one subject in Exp. 4 who were excluded from the experiments) consistently chose the banana pellet over the carrot piece on both trials.

2.2.2. Test trials

Apes participated in a total of 12 test trials split evenly between two testing sessions. Depending on the availability of animals, the delay between sessions was 1–11 days. Apes were confronted with the two transparent buckets containing the different populations consisting of banana pellets and carrots in each bucket. Items of the two types of food were of roughly equal size, but differed clearly in color and shape and could thus be easily distinguished. To aid the apes in gaining an overview of the two populations, the first experimenter (E1) shook the buckets several times and slightly tilted them forward. She then placed the buckets on the table and drew one item from each bucket (in such a way that the animal could not see which item it was because the drawing hand and the drawn object were occluded by other objects in the bucket), kept it invisibly in one hand, and then moved forward both hands simultaneously so that the ape could choose one of them. Apes chose an item by inserting their finger through one of the holes and touching the desired hand with the concealed food item inside. If the animal pointed to both hands simultaneously, the experimenter responded by saying: "Just one, [Name]" until the animal clearly chose a single hand. Apes then immediately received the food item as reinforcement. After the ape had made her decision, E1 handed her the chosen food item. The trial was over and E1 then removed the buckets from the table and out of the ape's sight so they could be refilled by E2 and placed them back on the table to start the next trial.

The side on which the more favorable population was positioned in a given session was counterbalanced across sessions and subjects. To rule out that low-level side preferences might suffice to solve the task (for the side with the more favorable population), E1 crossed here hands in half of the trials before offering the ape a choice. Trials with and without such crossing were administered in alternating order (it was counterbalanced across subjects which kind of trial came first) (see SI for details) (see Fig. 1).

2.3. Observational and coding procedure

For this and all following experiments, a second blind observer coded 25% of trials from video. Inter-rater reliability was excellent for all experiments ($\kappa > .86$).

2.4. Results

The mean proportion of trials in which apes chose the samples drawn from the two populations is depicted in Fig. 2. Apes as a group chose the hand from the bucket with the 4:1 distribution in 71% of the trials, significantly more often than expected by chance, $t(27) = 6.43$, $p = .001$ (Cohen's $d = .55$). (We detected no differences between the species, $F(3,24) = 1.58$, $p = .22$). This pattern cannot be due to learning over trials, as it was also reflected in trial 1 performance where 20 (71%) of the apes chose the hand from the 4:1 bucket, significantly more than expected by chance (Binomial test, $p = .02$; Cohen's $g = .21$).

3. Experiments 2 and 3: ruling out simple choice heuristics

3.1. Subjects

31 Apes (16 Chimpanzees, 4 Gorillas, 6 Orangutans and 5 Bonobos) were included in the final samples of Experiments 2 and 3. One further Chimpanzee was tested but failed to complete all trials due to lack of motivation.

3.2. Design and procedure

The general procedure was identical to Exp. 1 with the following exceptions: The aim of Experiments 2 and 3 was to rule out explanations in terms of superficial choice heuristics to the effect that apes' choices were simply based on a preference for buckets where the preferred kind of objects are in the majority. To do so, E1 sampled from two populations which had in common that banana pellets were in the minority (Exp. 2) or in the majority (Exp. 3), but which still differed in their frequency distributions (Exp. 2: 16:64 vs. 0:80/Exp. 3: 64:16 vs. 80:0).

3.3. Results

In Exp. 2, confronted with two buckets each containing 80 food items, with banana pellets: carrots distributions of 0:5 and 1:4, apes chose the latter in 84% of the trials, significantly more often than expected by chance, $t(30) = 11.42$, $p = .001$ (Cohen's $d = 1.06$), with a similar pattern on trial 1 where 87% of the apes chose from the 1:4 bucket, significantly more than expected by chance (Binomial test, $p = .001$, Cohen's $g = .37$). Our data revealed no differences between the species, $F(3,27) = 0.82$, $p = .49$).

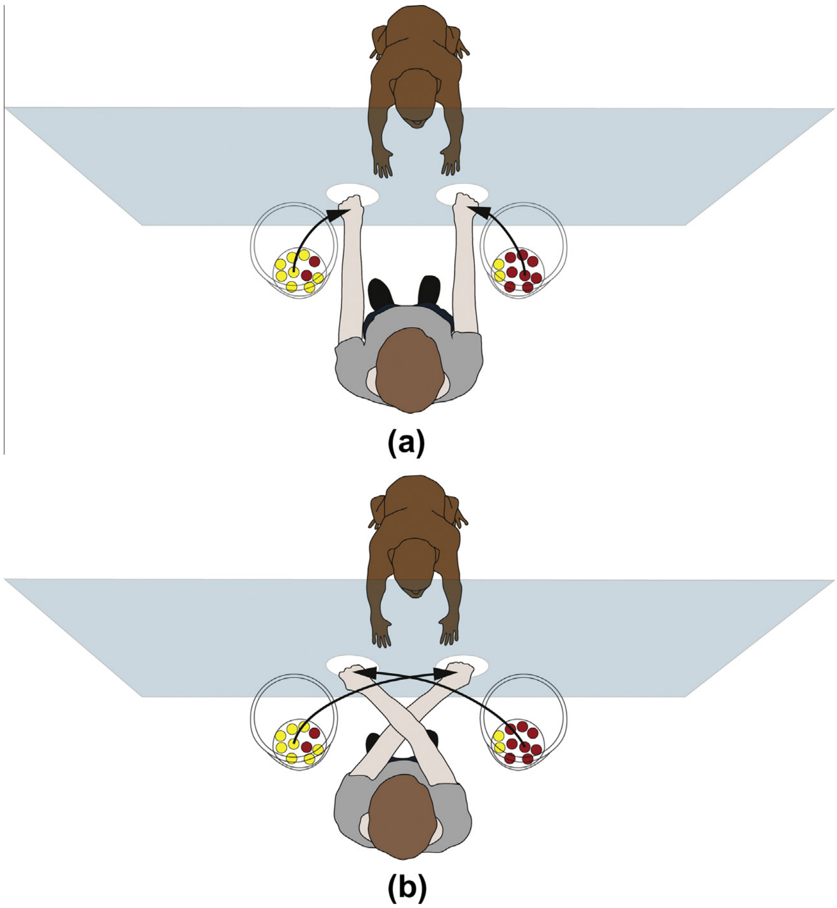

Fig. 1. Basic setup of the studies. An experimenter drew one item from each of the transparent buckets, kept the object invisibly in her hand, and then simultaneously moved both hands towards the ape, each behind a small hole in the plexiglas panel separating the ape and the experimenter. The ape then chose by inserting a finger through one of the holes. To rule out simple spatial heuristics (such as simply choosing the side where more attractive objects are), it was counterbalanced whether the hand with the object from one bucket was presented on the same side as the bucket (a) or on the opposite side (b).

In Exp. 3, confronted with two buckets with banana pellets: carrots distributions of 5:0 and 4:1, apes chose samples from the former in 62% of the trials, significantly more often than expected by chance, $t(30) = 3.84$, $p = .001$ (Cohen's $d = .36$),with a similar pattern on trial 1 where 61% of the apes chose from the 5:0 bucket (Binomial test, $p = .14$, Cohen's $g = .11$). One again, we detected no differences between the species, $F(3,27) = 1.40$, $p = .27$.

4. Experiment 5: ruling out Clever Hans effects

4.1. Subjects

26 Great apes (13 Chimpanzees, 3 Gorillas, 5 Orangutans and 5 Bonobos) were tested.

4.2. Design and procedure

Experiment 5 was designed to test whether the findings of Exp. 1 can be replicated when ruling out Clever Hans effects. To this end, any information (visual, tactile, auditory) about the items sampled on the part of the experimenter drawing and offering the objects was removed in the following way.

4.2.1. Blocking of visual access

Black cardboard was used to line the backside of buckets and to create visual occluders that were attached to either side of the top of the buckets to ensure that E1 had no visual access to the population distribution.

4.2.2. Blocking of tactile access

In order to prevent E1 from feeling from which population she was drawing, E1 actually moved her hand into a hidden compartment at the back of each bucket that was invisible to the apes. In order to ensure that the experimenter had no tactile information about the identity of each item sampled, she had two small plastic tubes attached to her palm which were pre-baited before each trial by E2. E1 then pretended to draw from the bucket, moving her hands into the hidden compartments in such a way that the tubes remained invisible to the subjects.

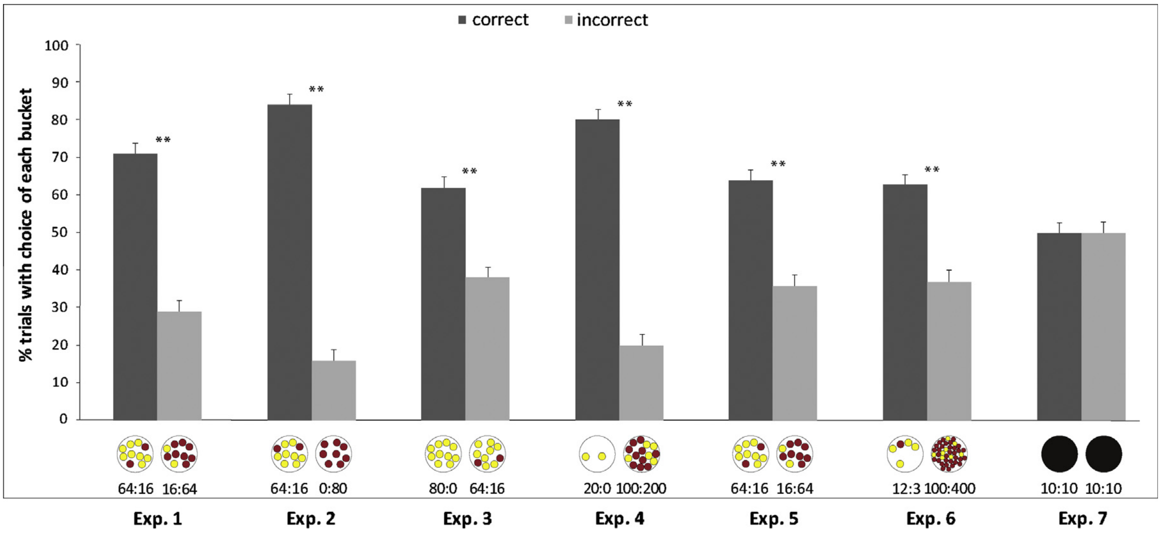

Fig. 2. Mean percentage of trials (with standard errors) in which the subject chose the correct/incorrect buckets. Below the graphs there are schematic representations of the distributions in the populations in both buckets in each study (yellow balls represent banana pellets (the preferred food items), orange balls represent carrots (less preferred). The ratios refer to the banana pellets: carrots ratio. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.3. Results

Controlling for Clever Hans effects, and with the same distributions of objects in the two buckets as in Experiment 1, the findings were replicated: apes ($N = 26$) chose the hand from the bucket with the 4:1 distribution in 64% of the trials, $t(25) = 4.84$, $p = .001$ (Cohen's $d = .52$), with a similar pattern on trial 1 where 69% of the apes chose from the 4:1 bucket (Binomial test, $p = .05$, Cohen's $g = .19$).

5. Experiments 4 and 6: do apes really represent relative (rather than absolute) frequencies?

5.1. Subjects

31 Apes (17 Chimpanzees, 3 Gorillas, 6 Orangutans and 5 Bonobos) were included in the final sample of Exp. 4. One further Gorilla was excluded from analysis due to inconsistency of item preference during the preference test. 26 great apes (13 Chimpanzees, 3 Gorillas, 5 Orangutans and 5 Bonobos) were tested in Exp. 6.

5.2. Design and procedure

Experiments 4 and 6 tested whether the inferences from populations to samples found in Exp. 1–3 were truly statistical inferences, based on relative frequencies, or whether they could be explained more parsimoniously in terms of representing absolute frequencies – absolute and relative frequencies were confounded in all but one previous infant studies (Denison & Xu, 2012) and in Exp. 1–3. In Exp. 4, therefore, apes had to choose from a bucket with 20 food items with a 20 banana pellet: 0 carrot distribution and a bucket with 300 food items and a 100 banana pellet: 200 carrots distribution. While the latter bucket contained a higher absolute number of banana pellets (100 vs. 20), what matters for statistical inferences from such a population to a sample is that the former was preferable in terms of the relative frequency of the desired items ($20/20 = 1$ vs. $100/300 = 0.33$).

Exp. 6 administered a similar design but included a Clever Hans control like Exp. 5 such that E1 pretended to sample from a secret hidden compartment with a special tube attached to her hand in such a way that she had no visual, tactile or other cues as to the identities of the populations or the samples. Two populations (buckets) containing items of each kind were used: apes had to choose from a bucket with 15 food items with a 12 banana pellets: 3 carrots distribution and a bucket with 500 food items and a 100 banana pellets: 400 carrots distribution, the latter bucket containing a higher absolute but lower relative number of banana pellets.

5.3. Results

In Exp. 4, apes chose the 20:0 bucket in 80% of the trials, significantly more often than expected by chance, $t(30) = 11.17$, $p = .001$ (Cohen's $d = 1.15$) (with a difference between species $F(3,27) = 3.212$, $p = .04$ due to the fact that all species but the gorillas performed above chance as a group). This pattern was also reflected in trial 1 performance where 22 (71%) of the apes chose the hand from the 20:0 bucket, significantly more than expected by chance (Binomial test, $p = .02$, Cohen's $g = .21$).

Similarly, in Exp. 6, apes chose the 12:3 bucket in 63% of the trials, significantly more often than expected by chance, $t(25) = 3.85$, $p = .001$ (Cohen's $d = .38$). This pattern was also reflected in trial 1 performance where 20 (77%) of the apes chose the hand from the 12:3 bucket, significantly more than expected by chance (Binomial test, $p = .01$, Cohen's $g = .27$).

6. Experiment 7: ruling out the use of olfactory cues

6.1. Subjects

20 Great apes were tested (7 Chimpanzees, 3 Gorillas, 5 Orangutans and 5 Bonobos).

6.2. Design and procedure

In order to test whether apes' choices in the previous experiments could have been based on olfactory information regarding which objects were in E1's hands, apes were confronted with E1 sampling banana pellets and carrots from two mixed populations (both with a banana pellet: carrot ration of 10:10) in two opaque buckets.

6.3. Results

Without any visual or other information about the populations drawn from, and with olfaction as the only source of information, apes chose the hand with the banana pellet in 50% of the trials, no different from chance, $M = 6$, $SD = 1.68$, $t(19) = 0$, $p = 1$.

7. Performance across experiments

Overall, apes performed above chance in all six studies which could be solved on the basis of intuitively statistical inferences (Exp. 1–6), but were at chance in Exp. 7 which could not be solved in this way. Across experiments, individual performance patterns were highly consistent. 24 apes participated in all six experiments in which the animals could choose correctly based on statistical information regarding the two populations randomly drawn from (Exp. 1–6). A Fisher's Omnibus Test exploring whether the distribution of Binomial test scores in the individual experiments was compatible with chance performance, showed that the vast majority of apes (20 out of 24) participating in all 6 studies performed above chance according to this test (see SI for details).

It is theoretically conceivable that apes solved each task by avoiding the sample drawn from the bucket with the higher absolute frequency of less desirable food items, not by taking into account relative frequencies. However, while this cannot be ruled out for each experiment taken by itself, we think that the results of all experiments taken together render this possibility unlikely. If apes' choices had been based on such avoidance they should have shown a greater avoidance for samples from those buckets with the higher absolute frequency of less desired items across experiments. This was not the case. For instance, there were 64 and 400 items of less desired food in the incorrect buckets (depicted on the right side in Fig. 2) of Experiments 5 and 6, respectively. However, subjects avoided both buckets at comparable levels (about 64% of the trials). Moreover, the incorrect bucket that received the greatest percentage of avoidance responses (about 84% of the trials) contained 80 items (Experiment 2), which is well below two other incorrect buckets that contained 200 (Exp. 4, 80% avoidance responses) and 400 (Exp. 6, 64% avoidance responses) less desired food items.

8. Discussion

The findings of the present experiments show that a basic form of drawing inferences from populations to samples is not uniquely human, but evolutionarily more ancient: It is shared by our closest living primate relatives, the great apes, and perhaps by other species in the primate lineage and beyond and it thus clearly antedates language and formal mathematical thinking both phylogenetically and ontogenetically.

8.1. How should the findings be best interpreted?

Yet, exactly what cognitive capacity do the present findings show? Do they reveal intuitive statistical reasoning properly so-called, that is, reasoning from relative frequencies (of favorable items in a given population relative to all items in the population) to predictions about random samples drawn from these populations? This is how adults would typically reason about such problems and this is how the cognitive capacities of infants recently amply documented in similar studies have been generally interpreted.

Or might these findings be explained more parsimoniously by simpler cognitive strategies and heuristics? The most obvious alternative would be that apes (and infants) might not reason about relative frequencies, but solve the tasks simply by discriminating absolute frequencies. And the most obvious and plausible version of this alternative would be that they discriminate the absolute frequencies of preferred items in each population. This alternative, however, can be ruled out empirically by the findings of the present Experiments 4 and 6 (explicitly designed for that purpose) for apes (and by similar recent control studies for infants; (Denison and Xu, 2013). A second possibility might be that apes engage in avoidance strategies involving the comparison of the absolute frequencies of dis-preferred items. Such a strategy, however, is not compatible with the present findings either, for the following reasons: If apes merely engaged in comparisons between the absolute frequencies of dis-preferred items, one would expect to find the discrimination functions and signature limits that is virtually always found in primates' (and humans') numerical discrimination of absolute set sizes exceeding the subitizing range (Cantlon, 2012). Discriminability of two populations should thus vary, following Weber's Law, as a function of the ratio of the absolute set sizes of dis-preferred items in each population. But this was clearly not the case in the present studies (for similar results regarding infants, see Denison and Xu (2013). In Experiments 1 and 5, for example, the ratio of carrot pieces in population 1 and population 2 was 1/4 whereas in Exp. 6 it was 3/400 – yet, despite these massive differences absolutely comparable rates of discrimination were found.

A third possibility might be that apes did not only focus on the absolute frequency of preferred items, nor on the absolute frequency of dis-preferred items but went beyond this and did take into account the relation of the two absolute frequencies within a given population to each other. However, instead of representing the crucial proportional relation (the relative frequency of preferred items relative to the whole population consisting of preferred and dis-preferred items), they might have used a much less complex heuristics based on difference scores (between the absolute frequencies of preferred and dis-preferred items). The simplest form of such an alternative might be a heuristic along the following lines: "If in a given population the frequency of preferred items is bigger than the frequency of dis-preferred items, choose samples from this population". This heuristics might explain some of the infant findings, and some of the results of the present study, but crucially it cannot explain the findings from Exp. 3 [64:16 vs. 80:0]. In this experiment, the frequency of preferred items is bigger than the frequency of dis-preferred items in both populations, and thus the heuristics would give the ape both the output "choose population 1" and the output "choose population 2"and should thus lead to chance behavior – yet apes still chose the population with the more favorable relative frequency of bananas.

Alternatively, the heuristic might be more subtle such that it is not only determined for each given population whether there are more preferred than dis-preferred items in that population, but the difference scores (between the absolute frequencies of preferred and dis-preferred items) might then be compared to each other. In other words, this would amount to a strategy of engaging in numerical discrimination of difference (rather than proportional) scores. Again, if this were the case, one would expect that this discrimination task would reveal the characteristics and signature limits found in virtually all numerical discrimination tasks with set sizes exceeding the subitizing range. Discriminability of populations should thus vary, following Weber's Law, as a function of the ratios of the difference scores to each other. But this was clearly not the case in the present study. Take, for example, Experiments 1 and 5, on the one hand, and Exp. 6 on the other hand. In all of the experiments, the ratio of the relative frequencies of preferred items to dis-preferred ones was kept constant (4/5 in population 1 vs. 1/5 in population 2), yet the differences scores between preferred and dis-preferred items varied. In Exp. 1 and 5 the difference scores are $64 - 16 = 48$ and $16 - 64 = -48$, whereas in Exp. 6 they are $12 - 3 = 9$ and $100 - 400 = -300$. Despite these massive differences in the relation of the difference scores to each other, however, these different experiments show absolutely comparable rates of discrimination across the experiments – and thus clearly suggest that apes tracked relative frequency and not the alternative differences scores. All in all, thus, the findings from the present experiments taken together are not compatible with any obvious simple heuristics but seem rather best explained by the assumption that apes are well capable of simple forms of intuitive statistical reasoning based on the representation of relative frequencies.

8.2. Implications for the comparative psychology of numerical cognition

Up to now, the numerical cognitive capacities known to be shared by humans and non-human primates comprised two systems for dealing with absolute set sizes – for the exact individuation of small sets (<4) and for approximate set size discrimination for arbitrarily large sets. The present findings are the first to show that beyond these two systems for representing absolute frequencies, we share with other apes the capacity to represent relative frequencies – a core foundation of statistical reasoning.

One question for future research concerns the relation of these capacities to each other: what roles do the systems for representing absolute set sizes play in the representation of relative frequencies both for small and for large sets? Relatedly, what are the properties and signature limits of the ability to distinguish relative frequencies? The system for approximate set size discrimination follows Weber's Law in humans and other primates (discriminability of two sets depends on the ratio of the absolute set sizes: if the system can discriminate 5 bananas from 10 bananas, it can discriminate 10 from 20, 30 from 60, etc. (Cantlon & Brannon, 2007; Xu & Spelke, 2000). And recent research with human children (McCrink & Wynn, 2007; Sophian, 2000) and non-human animals (Emmerton, 2001; Wilson, Britton, & Franks, 2002; Woodruff & Premack, 1981) as well as work in cognitive neuroscience (Jacob, Vallentin, & Nieder, 2012; Vallentin & Nieder, 2008; Yang & Shadlen, 2007) suggests that discrimination of ratios might be subject to the same signature limit in accordance with Weber's Law (discrimination breaks down as a function of the ratio of the ratios to be discriminated) as the discrimination of sets (which breaks down as a function of the ratio of the absolute set sizes). Future studies will thus need to test whether intuitive statistical reasoning behaves similarly one level up, by discriminating any two relative frequencies of as a function of the ratio of the relative frequencies to each other. If such a pattern were found, this would be first evidence to suggest that intuitive statistics might be based on similar or the same processes as approximate absolute set size discrimination.

Another question concerns the relation of the capacity for intuitive statistics as documented here –inferentially relating populations and randomly drawn samples- to the capacity to represent statistical information found in many species in domains such as auditory pattern extraction (Abe & Watanabe, 2011; Fitch & Hauser, 2004; Gentner, Fenn, Margoliash, & Nusbaum, 2006; Toro & Trobalón, 2005), risk assessment and decision making (Balci et al., 2009) or optimal foraging (Stephens, 2008). Are all of these phenomena manifestations of a common underlying domain-general capacity to deal with information regarding distributions and relative frequencies? Or are they separate and fragmented, perhaps modular capacities with little inter-connections? Auditory statistical pattern extraction, for example, arguably is such a special and potentially separate domain (possibly having to do with learning the "grammar" of songs or other communicative systems) that it is highly unclear what its relation is to a more general capacity for flexibly drawing inferences from populations to samples and vice versa.

A broader question, finally, concerns the relation of such intuitive statistics to other kinds of reasoning: In humans, statistical information is systematically integrated with other types of information from very early on: Even infants, for example. understand that mechanical constraints (e.g. only some kinds of objects in a population a physically drawable) or psychological factors (the person drawing likes one kind of item more) can turn a sampling process into a non-random one such that the sample need not reflect the distribution in the population (Denison & Xu, 2010a; Téglás et al., 2007, 2011). Whether such systematic integration of different cognitive domains can be found in non-human animals is an exciting open question with potentially far-reaching theoretical ramifications: some influential theories of comparative cognition view this very cross-domain integration as one of the hallmarks of uniquely human cognition (Carruthers, 2002; Gopnik & Schulz, 2004; Penn, Holyoak, & Povinelli, 2008; Spelke, 2003; Woodward, 2007).

Author contributions

H.R., A.C. and J.C designed the experiments, L.S., N.S., A.G. and J.M. ran the experiments. All authors analyzed the results and discussed the findings. H.R., A.C. and J.C wrote the paper.

Acknowledgements

We would like to thank Charlotte Rahn, Roger Mundry, Daniel Hanus, Hanna Petschauer, Johannes Grossmann, Kim Gärtner and all Zoo keepers involved for their advice and help at various stages of this project. Thank you very much to Marike Schreiber for preparing the graphs. We would like to thank Daniel Haun, Tamar Kushnir, Jutta Müller, Marco Schmidt, Claudio Tennie and Michael Tomasello for comments on an earlier draft of this paper.

This work was supported by a "Dilthey Fellowship" of the Volkswagen Foundation and the Fritz Thyssen Foundation and by the German Initiative of Excellence.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2013.12.011.

References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience, 14*(8), 1067–1074.

Balci, F., Freestone, D., & Gallistel, C. R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Sciences, 106*(7), 2459–2463. http://dx.doi.org/10.1073/pnas.0812709106.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science, 282*(5389), 746–749.

Cantlon, J. F. (2012). Math, monkeys, and the developing brain. *Proceedings of the National Academy of Sciences, 109*(Supplement 1), 10725–10732. http://dx.doi.org/10.1073/pnas.1201893109.

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

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

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

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences, 25*(6), 657–726.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*(1), 1–73.

Denison, S., & Xu, F. (2012). Single-event probabilistic inference guides 10- to 13-month-old infants' choice behavior. Paper presented at the International Conference on Infancy Studies, Minneapolis.

Denison, S., & Xu, F. (2013). The origins of probabilistic inference in human infants. *Cognition, 130*(3), 335–347. http://dx.doi.org/10.1016/j.cognition.2013.12.001.

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

Denison, S., & Xu, F. (2010b). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science, 13*(5), 798–803.

Emmerton, J. (2001). Pigeons' discrimination of color proportion in computer-generated visual displays. *Animal Learning and Behavior, 29*(1), 21–35. http://dx.doi.org/10.3758/bf03192813.

Fitch, W., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*(5656), 377–380.

Flombaum, J. I., Junge, J. A., & Hauser, M. D. (2005). Rhesus monkeys (*Macaca mulatto*) spontaneously compute addition operations over large numbers. *Cognition, 97*(3), 315–325.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*, 1204–1207.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*(4), 684–704.

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences, 8*(8), 371–377.

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

Hauser, M., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceeding of the National Academy of Science of the USA, 9*(4), 1514–1517.

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

Kamil, A. C., Krebs, J. R., & Pulliam, H. R. (1987). *Foraging behavior*. New York: Plenum Press.

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

Ma, L., & Xu, F. (2011). Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition, 120*(3), 403–411.

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

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

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. (Trans L. Leake et al.).

Sophian, C. (2000). Perceptions of proportionality in young children: Matching spatial ratios. *Cognition, 75*(2), 145–170. http://dx.doi.org/10.1016/s0010-0277(00)00062-7.

Spelke, E. (2003). What makes us smart? In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind. Advances in the study of language and thought* (pp. 277–311). Cambridge, MA: MIT Press.

Stephens, D. W. (2008). Decision ecology: Foraging and the ecology of animal decision making. *Cognitive, Affective, and Behavioral Neuroscience, 8*(4), 475–484.

Stüttgen, M., Yildiz, A., & Güntürkün, O. (2011). Adaptive criterion setting in perceptual decision making. *Journal of the Experimental Analysis of Behavior, 9*(2), 155–176.

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

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

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Attention, Perception, and Psychophysics, 67*(5), 867–875.

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

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

Vallentin, D., & Nieder, A. (2008). Behavioral and prefrontal representation of spatial proportions in the monkey. *Current Biology, 18*(18), 1420–1425. http://dx.doi.org/10.1016/j.cub.2008.08.042.

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

Woodruff, G., & Premack, D. (1981). Primative mathematical concepts in the chimpanzee: Proportionality and numerosity. *Nature, 293*(5833), 568–570. http://dx.doi.org/10.1038/293568a0.

Woodward, J. (2007). Interventionist theories of causation in psychological perspective. In A. Gopnik & L. Schulz (Eds.), *Causal Learning: Psychology, Philosophy and Computation* (pp. 19–36). Oxford: Oxford University Press.

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

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11.

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature, 447*(7148), 1075–1080. http://dx.doi.org/10.1038/nature05852.