ABC 2021, 8(3):305-324
Animal Behavior and Cognition
DOI: https://doi.org/10.26451/abc.08.03.01.2021

The Ape Lottery: Chimpanzees Fail To Consider Spatial Information When Drawing Statistical Inferences

Johanna Eckert[1,2,3,*], Hannes Rakoczy[2], Shona Duguid[3,4], Esther Herrmann[3,5], and Josep Call[3,4]

[1]Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[2]Department of Developmental Psychology, University of Goettingen, Germany
[3]Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig Germany
[4]School of Psychology and Neuroscience, University of St Andrews, United Kingdom
[5]Department of Psychology, University of California Berkeley, USA

*Corresponding author (Email: johanna_eckert@eva.mpg.de)

Citation – Eckert, J., Rakoczy, H., Duguid, S., Herrmann, E., & Call, J. (2021). The ape lottery: Chimpanzees fail to consider spatial information when drawing statistical inferences. *Animal Behavior and Cognition, 8*(*3*), 305-324. https://doi.org/10.26451/abc.08.03.01.2021

Abstract – Humans and nonhuman great apes share a sense for intuitive statistics, making intuitive probability judgments based on proportional information. This ability is of tremendous importance, in particular for predicting the outcome of events using prior information and for inferring general regularities from limited numbers of observations. Already in infancy, humans functionally integrate intuitive statistics with other cognitive domains, rendering this type of reasoning a powerful tool to make rational decisions in a variety of contexts. Recent research suggests that chimpanzees are capable of one type of such cross-domain integration: The integration of statistical and social information. Here, we investigated whether apes can also integrate physical information into their statistical inferences. We tested 14 sanctuary-living chimpanzees in a new task setup consisting of two "gumball machine"-apparatuses that were filled with different combinations of preferred and non-preferred food items. In four test conditions, subjects decided which of two apparatuses they wanted to operate to receive a random sample, while we varied both the proportional composition of the food items as well as their spatial configuration above and below a barrier. To receive the more favorable sample, apes needed to integrate proportional and spatial information. Chimpanzees succeeded in conditions in which we provided them either with proportional information or spatial information, but they failed to correctly integrate both types of information when they were in conflict. Whether these limitations in chimpanzees' performance reflect true limits of cognitive competence or merely performance limitations due to accessory task demands is still an open question.

Keywords – Intuitive statistics, Probabilistic reasoning, Physical cognition, Cross-domain integration, Primates, Great apes

Intuitive statistical reasoning is the capacity to make intuitive probabilistic inferences based on relations between populations, sampling procedures and resulting samples. This capacity is indispensable to our daily lives and one of the hallmarks of human thinking. We constantly use sample observations to draw general inferences about the world, to foresee the outcome of events, and to make rational decisions under uncertainty. Despite the ubiquity of situations presumably requiring statistical inference, historically, statistical reasoning was thought to develop late in ontogeny (Piaget & Inhelder, 1975), to be biased by general-purpose heuristics throughout adulthood (Ayton & Fischer, 2004; Chapman & Chapman, 1969; Kahneman & Tversky, 1972, 1973; Nisbett & Ross, 1980; Tversky & Kahneman, 1974, 1983), and to be restricted to specific situations and types of information (Cosmides & Tooby, 1996; Gigerenzer, 1991; Gigerenzer & Hoffrage, 1995; McDowell & Jacobs, 2017) .

In the last fifteen years, however, due to the emergence of new non-verbal procedures, evidence has accumulated from developmental research showing that even pre-verbal infants can reason from populations of items to randomly drawn samples and vice versa (Denison et al., 2013; Denison & Xu, 2010a, 2014; Teglas et al., 2007, 2015; Xu & Garcia, 2008). For instance, when infants were presented with two transparent jars containing mixtures of pink (preferred) and black (non-preferred) lollipops, they intuitively chose a random (covered) sample drawn from the jar with a greater proportion of pink to black lollipops, which was thus more likely to be a preferred candy (Denison & Xu, 2010a).

Crucially, intuitive statistics is not a cognitively isolated process; from infancy onwards, humans are able to integrate a variety of different types of information into their statistical inferences to make rational judgments (Denison et al., 2014; Denison & Xu, 2010b; Gweon et al., 2010; Lawson & Rakison, 2013; Teglas et al., 2007, 2011; Wellman et al., 2016; Xu & Denison, 2009). For example, pre-verbal infants already make use of their knowledge of the social world when judging whether a sampling process will be random or not (Gweon et al., 2010; Wellman et al., 2016; Xu & Denison, 2009). When infants faced an experimenter who had previously expressed a preference for one type of object, infants expected the sample to reflect these preferences, regardless of the populations' composition. Importantly, they only did so when the experimenter had visual access to the population while sampling. When the same biased experimenter sampled blindly, infants expected the sample to be of the majority type of the population (Xu & Denison, 2009). These findings demonstrated that infants flexibly considered intuitive psychological knowledge to judge the sampling conditions and drew according statistical inferences.

Results paralleling those of the social domain were also found for the physical domain (Denison et al., 2014; Denison & Xu, 2010b; Lawson & Rakison, 2013; Teglas et al., 2007, 2011). One study (Teglas et al., 2007), for instance, used the violation of expectation looking time paradigm. This paradigm utilizes the fact that infants usually look longer at scenes that violate their expectations (i.e., events they find unlikely or impossible). In this study, infants were first familiarized with a scene of four bouncing objects in a lottery machine: three of them were yellow, and one was blue. Subsequently, after a short occlusion phase, infants watched one of the objects exiting the lottery machine: either one of the yellow objects, or the blue object. Infants looked longer at the unlikely outcome of a minority object exiting the lottery machine. In a second experiment, a horizontal barrier was inserted into the lottery machine, separating the two object types. The three yellow objects were above the barrier, and therefore were prevented from exiting the lottery machine, whereas the single blue object was below the barrier, and could still exit. In this case, infants´ looking times showed the reverse pattern, indicating that they were not surprised to see the minority object exiting when it was the only possible outcome. In a modified version of the same paradigm, Teglas and colleagues (2011) tested infants for their ability to integrate spatiotemporal information with statistical inference. Again, the lottery machine contained three yellow objects and one blue object, all of them bouncing in random patterns. This time the authors varied the spatial arrangement of the items immediately before an occlusion phase (i.e., either the single blue object or one of the three yellow objects was close to the opening), as well as the occlusion duration. Here, infants´ looking times followed a graded pattern: when the occlusion duration was very short, infants seemed to judge the situation based on the spatial arrangement immediately before the occlusion and expected that object closest to the opening to exit, regardless of whether it was of the minority or majority type. When occlusion lasted longer, infants ignored the spatial arrangement prior to occlusion, and expected one of the majority objects to exit. When occlusion duration was intermediate, infants´ looking times were intermediate, too, and thus reflected both the object proportions and their distance from the opening. Hence, infants integrated information about the ratio of objects, their physical arrangement and occlusion time to judge the outcome of an event.

Together, these studies demonstrated that pre-verbal infants not only reason from population to sample, but they also take into account social and physical variables when drawing statistical inferences. These findings suggest that humans can integrate substantive domain-knowledge into their probabilistic inference mechanism from infancy onwards. The ability to transfer and combine information across different domains is, aside from context and stimulus independence, one source of evidence for domain-generality (Burkart et al., 2017; Carruthers, 2002; Gentner et al., 2001; Mithen, 1996, also see Hermer &

Spelke, 1996; Spelke, 2003). Presumably, a domain-general statistical inference mechanism is a powerful tool for infants that guides their learning and helps them acquire their rapidly growing knowledge about the world.

The fact that sophisticated statistical reasoning abilities require neither formal education nor language was further confirmed by recent comparative research: Rakoczy and colleagues (2014) presented chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), orangutans (*Pongo abelii*), and gorillas (*Gorilla gorilla*) with a task similar to one previously used in developmental research (Denison & Xu, 2010a). In a series of seven experiments, the apes were confronted with two transparent buckets containing mixed populations of preferred and non-preferred food items (banana pellets and carrot pieces) in specific ratios. Subsequently, the experimenter randomly drew one sample from each in a way that the subject could not see what was drawn. Then the subject was given a choice between the two covered samples. To receive a preferred food item (pellet), apes had to discriminate between the two populations with regards to their proportions of pellets to carrots and form corresponding expectations about the probability of sampling a pellet from each of them. Apes were able to infer which of the two populations was more likely to lead to a pellet as a sample across conditions, and this was apparent from the very first trial onwards. Moreover, they chose correctly even when absolute and relative frequencies were disentangled (i.e., when the population with the more favorable ratio of pellets to carrots contained absolutely fewer pellets than the other one). Subsequent studies confirmed these findings (Eckert, Call et al., 2018) and demonstrated that, under some circumstances, apes make reverse inferences: reasoning from sample to population (Eckert et al., 2017). Species of both Old and New World monkeys did not succeed in all conditions of a comparable task involving inferences from population to sample (long-tailed macaques, *Macaca fascicularis*: Placì et al., 2018; capuchin monkeys, *Sapajus* spp.: Tecwyn et al., 2017). However, more recent research using different paradigms did find some evidence for basic types of probabilistic inference in two Old World monkey species (rhesus macaques, *Macaca mulatta*: De Petrillo & Rosati, 2019; long-tailed macaques: Placì et al., 2019) and even in two species of parrots and pigeons (Kea, *Nestor notabilis*: Bastos & Taylor, 2020; a Grey parrot, *Psittacus erithacus*: Clements et al., 2018; White King pigeons, *Columba livia domestica*: Roberts et al., 2018). Hence, intuitive statistics is not a uniquely human capacity, but appears to be part of our evolutionary heritage and was probably already present in the last common ancestor of humans and other apes, and perhaps even earlier in evolutionary history.

Whereas the existence of intuitive statistical abilities in great apes is well documented, we hardly know anything about whether non-human animals can integrate statistical inference with other cognitive capacities. For non-human primates, so far only one study tested for one type of such cross-domain integration: Eckert, Rakoczy and colleagues (2018) showed that chimpanzees, just like human infants, consider an experimenter's mental states when drawing statistical inferences. In this study, chimpanzees were tested in the previously established "bucket paradigm" (Eckert, Call et al., 2018; Rakoczy et al., 2014) that required them to infer which of two mixed populations of preferred and non-preferred food items was more likely to lead to a desired outcome for the subject. Through several experiments, the authors manipulated whether experimenters had preferences for drawing certain objects (in comparison to random drawing) and whether they had visual access to the population while sampling or drew blindly (following the task procedure previously used in developmental research; Xu & Denison, 2009). Results suggested that chimpanzees assumed random sampling which reflected the population's distribution when they had no prior information about the experimenters who sampled. If, however, the apes had reason to assume that the experimenters were biased, subjects' choices reflected these biases. The extent of this influence was dependent on whether the experimenters had visual access to the populations or not. Hence, this study was the first to demonstrate that chimpanzees were able to flexibly integrate two sources of information to make rational decisions under uncertainty, and thereby provided the first evidence that apes' statistical inference mechanism also may be domain-general.

Recent research suggests that such capacities are not unique to primates, or even to mammals, but may have emerged several times throughout evolutionary history in analogous ways. Bastos and Taylor (2020) found that a small sample of kea, a parrot species endemic to New Zealand, reasoned from

population to sample in a task based on the ape study (Rakoczy et al., 2014). They also seemed to consider information about an experimenter's choice biases similarly to chimpanzees (but note that the birds were tested in a very simplified version of this task). Crucially, kea also considered spatial information when drawing statistical inferences. The birds were presented with two jars, which both had a horizontal barrier placed in their center. Both jars contained identical overall populations of rewarding and non-rewarding tokens, but the proportions differed above and below the barriers. Kea only paid attention to the part of the population located above the barrier when choosing between randomly drawn samples, disregarding the inaccessible section of the population. Hence, the birds seemingly integrated knowledge of a physical barrier into their predictions of a sampling outcome. While more research is needed to confirm these first findings, the results do suggest that domain-general statistical thought may have evolved in convergent manner, leading to analogous abilities in distantly related species.

An important open question is whether great apes, like human infants and kea, can also integrate physical information when drawing statistical inferences. If that was the case, it would suggest that great apes, like humans, possess a domain-general inference mechanism, and hence an efficient tool allowing them to rapidly acquire knowledge about their environment by drawing general conclusions from sparse data and to use these generalizations to predict the outcome of events in a variety of different contexts.

The aim of the current study was to investigate whether apes can combine their intuitive sense of statistics with their intuitive sense of physics in a new task setup. While previous research has shown that apes have a sense of intuitive statistics as well as a sense of intuitive physics (e.g., they respond appropriately in tasks requiring a basic understanding of gravity and solidity; Cacchione & Call, 2010; Cacchione et al., 2009; Cacchione & Krist, 2004), it remains unclear whether chimpanzees can combine these two sources of information and integrate knowledge of a physical constraint into their statistical inferences. We used a novel task setup (the "ape lottery"), in which chimpanzees acted as sampling agents by drawing food items from two gumball machine-like apparatuses. In each condition, the subject could decide which of two apparatuses they wanted to operate to receive a random sample, while we varied both the proportional composition of the food items in the apparatuses as well as their spatial distribution above and below a horizontal barrier.

Methods

Subjects

Sixteen chimpanzees (N_{female} = 11) between 10 and 33 years of age participated in the familiarization phase of the current study. One male did not pass the familiarization criterion (see below) and one female chose not to enter the testing enclosure during the time of testing. Hence, fourteen chimpanzees took part in the test conditions (see Table A1 for subject details). All individuals were born in the wild and were orphaned at a young age before they were transferred to Ngamba Island Chimpanzee Sanctuary, Uganda. At the time of testing, all chimpanzees lived in a social group of 49 individuals. In accordance with the recommendations of the Weatherall report 'The use of nonhuman primates in research' subjects were roaming freely on the 40 ha island covered with tropical rainforest during the day and spent the night in seven interconnected sleeping rooms (approximately 140 m^2) with water ad libitum and regular feedings. Subjects were never food or water deprived and participated in the study on a voluntary basis. Animal husbandry and research comply with the 'PASA Primate Veterinary Healthcare Manual' and the 'Guidelines for the Treatment of Animals in Behavioral Research and Teaching' of the Association for the Study of Animal Behavior.

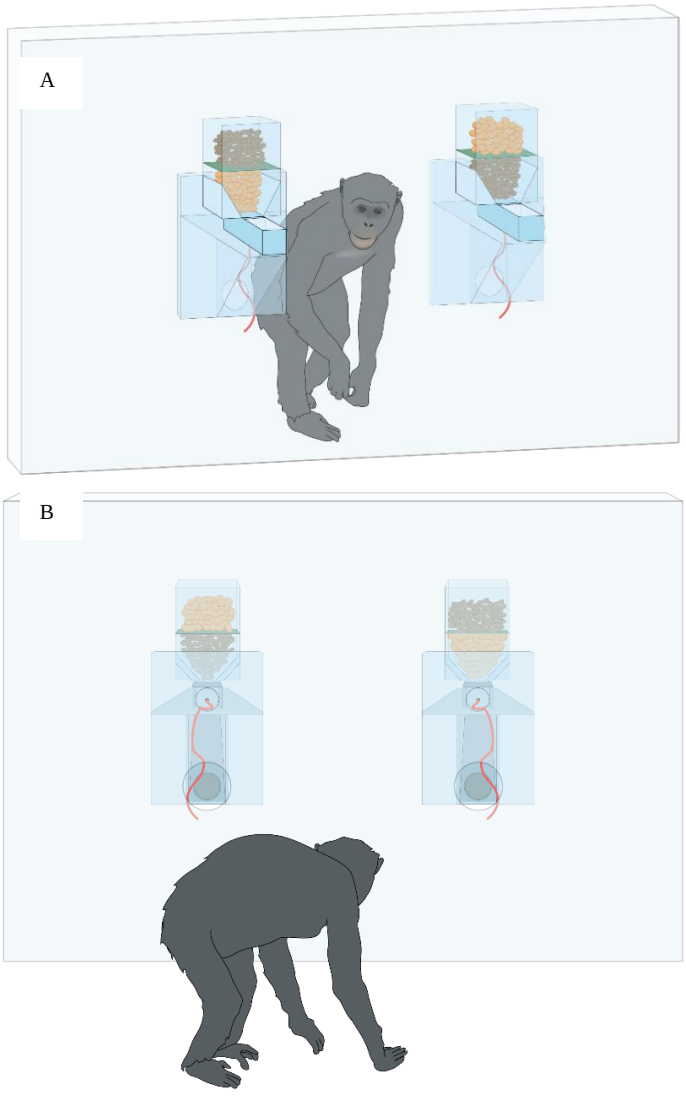
Apparatus and Materials

The setup consisted of two identical apparatuses (height: 40 cm, width: 35cm, length: 25 cm) made of plexiglass that were mounted on the testing room mesh from the outside. The upper part of each

apparatus consisted of a removable transparent container (height: 25 cm, width: 6 cm, length: 20 cm) that could be filled with populations of food items. The container had a quadratic opening (5 x 5 cm) on the bottom. Underneath the container was a drawer (height: 5 cm, width: 7 cm, length: 15 cm) that could be pulled by a string (length: 25 cm). While the front section of the drawer was a complete cuboid, the back section was a cuboid with openings both at the top and the bottom. Once the experimenter placed the string in reach of the subject, it could pull the string and move the drawer so that the opening aligned with the opening on the bottom of the container. This caused a food item to fall out of the container into a transparent tunnel underneath the drawer (height: 25 cm, width: 9 cm, length: 41 cm). The subject could access this tunnel and remove the food item. During test conditions, the apparatuses were mounted on the testing room mesh 50 cm apart (see Figure 1 A and B, as well as these links to Video 1a and Video 1b for an illustration of the setup and the apparatus).

Figure 1

Experimental Setup from the Experimenter's Perspective (A) and the Chimpanzee's Perspective (B)



Note. The two apparatuses could be filled with populations of peanuts and carrots. The subject was allowed to choose which of the two apparatuses to operate by pulling a string attached to the lower part of it. By doing so, one food item was released from the bottom of the machine.

For the experimenter to have control over the type of food item being released from the population, the opening of the removable container could be sealed by the means of transparent tape. In addition, the apparatus had a hidden compartment behind the drawer, which could be pre-baited with a food item. We ensured that subjects did not see this pre-baited compartment by attaching opaque tape to the front and side of it, as well as on the lower 2 cm of the container. For some test conditions, a green barrier (height: 2 mm, width: 5 cm, length: 19 cm) could be inserted horizontally into the container, leading to a spatial segregation of the food items in the apparatus (a string attached to the barrier facilitated its placement in and removal from the container).

We used two types of food: peanuts (the preferred food type) and pieces of carrots (the non-preferred food type). The stable preference hierarchy among these two food types was confirmed in several previous studies with the same individuals (e.g., Eckert, Call et al., 2018; Eckert, Rakoczy et al., 2018). Only those subjects who continued showing this preference in the current study passed the familiarization phase and proceeded to the actual test conditions (see below).

Design and Procedure

Familiarization

Subjects were familiarized with the apparatus and the setup in two steps: In the first step, a single apparatus was mounted on the mesh. The subject witnessed it being baited with one food item (in half of the trials a peanut, in the other half of trials a piece of carrot). Subsequently, the experimenter put the string in reach of the subject who was then allowed to pull it and thereby release the food item. Importantly, in this first step the apparatus was entirely transparent in order to ensure that subjects were able to see the full workings of the apparatus. Subjects received 12 trials presented in randomized order in a single session. All subjects managed to retrieve the food items in all 12 trials. In a second step, subjects were presented with the two apparatuses mounted at a distance of 3 m. This time the drawer part of both apparatuses, as well as the lower 2 cm of the containers were occluded by opaque tape. During the first four trials of this type of familiarization, just one of the two apparatuses was baited with 100 pieces of food items (only peanuts in half of the trials, only carrot pieces in the other half of the trials). The order of peanut and carrot trials was alternated. Each population was presented once on each side. Then the subject was allowed to pull the string and release one food item. In the subsequent eight trials, both apparatuses were baited: one with 100 peanuts, the other with 100 pieces of carrot (the side of the peanut container was counterbalanced, and the order of trials was randomized). The experimenters placed the strings of both apparatuses in reach of the subject. The subject was then allowed to choose between them and release one food item out of one of the two apparatuses. The next trial began immediately after the food item was retrieved. Only those subjects who reliably chose the apparatus containing peanuts in at least 7 out of 8 trials proceeded to the subsequent test phase. Subjects who did not fulfill this criterion (N = 7) received another session with 12 trials with the two baited apparatuses on the next day. Five of seven subjects reached criterion (choosing the peanut population in at least 10 out of 12 trials) after this second session. Prior to the test conditions, all subjects were handed the barrier and were allowed to inspect it as long as they wanted to ensure that they were aware of the solidness of the material.

Pilot test

Prior to conducting the test conditions, we administered a slightly different version of the same test: in this test the apparatuses were mounted at a distance of 3 m (as in the familiarization phase), and each apparatus was presented to the subjects by one of two experimenters (see Supplementary Material for more detailed information). In this version of the study, chimpanzees performed at (or below) chance level across conditions. According to the experimenters´ observations, the chimpanzees in this study did not pay attention to the populations within the containers, but rather focused on the experimenters, who seemed to be the more salient stimuli to the apes. In order to circumvent this interference, we repeated the

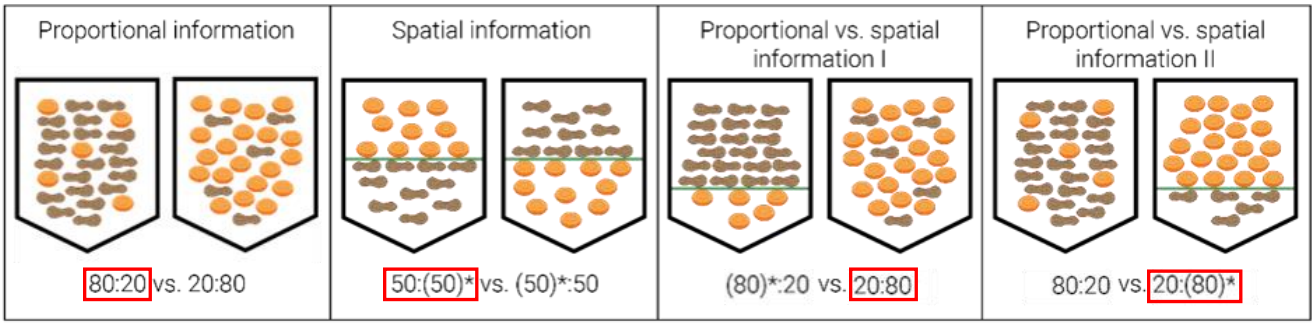
tests with both apparatuses mounted closer together, which allowed us to reduce the number of experimenters to one. Please see the Supplementary Material for more detailed information on the procedure and results of this prior version of the study.

Test conditions

Chimpanzees were tested in four conditions (see Figure 2 for a schematic illustration of the test conditions). The order of conditions was randomized for each subject. Each condition was tested in a single session with twelve trials. Similar to the second part of the familiarization, the contents of the drawer of the apparatus were concealed by opaque tape. Before a test trial started, the drawer of each apparatus was pre-baited with one food item outside of the subject´s view. Then, subjects witnessed the container of both apparatuses being baited with mixtures of peanuts and carrots. There were always 100 food items in total. This was a 4:1 mixture in one apparatus, and a 1:4 mixture in the other, except for the "Spatial information" condition where the distribution was 1:1 in both apparatuses. In conditions in which one or both apparatuses contained a barrier, the baiting was stepwise: first, one type of food was filled in, then the barrier was positioned, then the second type of food was added. The side of the more favorable population was counterbalanced. Subsequently, the experimenter placed the strings of both apparatuses in reach of the subject who was then allowed to choose between the two apparatuses and pull the string of one of them. A trial ended once a food item was released (see Video 2 for an example of a test trial).

Figure 2

Schematic Illustration of the Test Conditions



Note. *Numbers in parentheses indicate inaccessible proportion of food. Red squares indicate the more favorable option in each condition. Subjects were tested in four test conditions in counterbalanced order, in which we varied both the proportional composition of preferred (peanuts) and non-preferred food items (carrot pieces), as well as their spatial arrangement. In each condition, subjects were allowed to choose which of the two apparatuses to operate in order to receive a sample drawn from the bottom of the apparatus.

Proportional Information. This condition investigated whether chimpanzees were able to use proportional information in order to choose the apparatus which was more likely to lead to a preferred food item as a randomly drawn sample. The containers held mixed populations of peanuts and carrot pieces with a random spatial distribution. While one population contained a proportionally more favorable population (80 peanuts and 20 pieces of carrots: 80% chance of a peanut sample) the other contained a less favorable population (20 peanuts and 80 pieces of carrots: 20% chance of a peanut sample; see Figure 2 for an illustration of the two distributions). To facilitate comparability with previous studies (Eckert, Call et al., 2018; Eckert, Rakoczy et al., 2018; Rakoczy et al., 2014), the (seemingly) randomly drawn samples were always of the majority type (i.e., the food in the higher proportion was always the one delivered from that apparatus).

Spatial information. In this condition, we investigated whether apes inferred that a physical constraint (the spatial distribution of food items) would determine the outcome of the sampling event.

More specifically, while both apparatuses contained the same proportions of peanuts to carrot pieces (50:50), the two food types were segregated by a barrier in contrasting ways: for the more favorable population (100% chance of a peanut sample), all peanuts were under the barrier, close to the opening of the container, whereas the carrot pieces were located above the barrier, separated from the opening. For the less favorable population (0% chance of a peanut sample), the spatial arrangement was the opposite, with the carrot pieces close to the opening and the peanuts separated by a barrier (see Figure 2 for an illustration of the distributions). Here, the sampled food items were always of the food type located underneath the barrier.

Proportional vs. Spatial Information I. In this condition, as well as in the "Proportional vs. spatial information II" condition (see below), we explored whether chimpanzees could integrate proportional and spatial information when both were in conflict. More precisely, one population contained a more favorable proportion of peanuts to carrots (80:20), but the food types were segregated in a way that all peanuts were located above the barrier – thus far away from the container´s opening, rendering the chance to sample a peanut at 0%. The other population contained a less favorable proportion of peanuts to carrots (20:80), but the two food types were intermixed, leaving a 20% chance of sampling a peanut (see Figure 2 for an illustration of the distributions). To avoid frustration and maintain the apes' motivation, the rewarding scheme reflected the proportions that could be inferred from the apparatuses as closely as possible: in the latter population, half of the subjects had the chance of obtaining a peanut in two out of twelve trials (i.e., in 16,6% of trials), the other half had the chance of obtaining a peanut in three out of twelve trials (i.e., in 25% of trials). Such peanut trials were interspersed pseudo-randomly among all trials (with at least one carrot trial between two peanut trials). If chimpanzees understood that in the 80:20 population (but not in the 20:80 population) it was spatial arrangement, rather than proportional composition, that determined the outcome, they should choose to pull the string of the apparatus with the intermixed 20:80 population.

Proportional vs. spatial information II. Here, in the container holding the proportionally less favorable population (20:80) the two food types were segregated in a way that all peanuts were located underneath the barrier, close to the exit, leading to a 100% chance of sampling a peanut from this apparatus. In the proportionally more favorable (80:20) population the food types were intermixed, rendering chances for a peanut sample 80% (see Figure 2 for an illustration of the distributions). Again, the rewarding scheme corresponded to the objective chances of obtaining each food item from each apparatus as closely as possible. If subjects inferred that spatial arrangement better indicated reward type than proportional composition in the 20:80 population but not in the 80:20 population, they should choose the apparatus holding the segregated 20:80 population.

Analyses

We coded subjects' choice between populations (correct vs. incorrect) and analyzed the data using Generalized Linear Mixed Models (GLMM; Baayen, 2008) with binomial error structure. As fixed effects we included condition, session number (where each session was one of the four conditions) and trial number (to check for potential learning effects), as well as the three-way-interaction between condition, session number and trial number. To control for a potential effect of subjects' age, we included age and age^2 (to control for potential non-linear effects) as further fixed effects. Subject ID was included as random effect. To keep type I error rate at the nominal level of 5% (Barr, 2013; Schielzeth & Forstmeier, 2009) we included all possible random slopes components (condition, session number, and trial number) and the respective correlations between random slopes and intercepts. Session number, trial number, age and age^2 were z-transformed (to a mean of zero and a standard deviation of one). Variance Inflation Factors (VIF; Field, 2013) were derived for a standard linear model excluding the random effects and interactions, using the function vif of the R-package car (Fox & Weisberg, 2018) and did not indicate collinearity to be an issue. We assessed model stability by comparing the estimates derived by a model based on all data with those obtained from models with the levels of the random effects excluded one at a time. This revealed that the model was stable. The significance of the full model as compared to

the null model (comprising only age, age^2 and the random effect subject ID) was established using a likelihood ratio test (R function ANOVA with argument test set to "Chisq"; Dobson & Barnett, 2008; Forstmeier & Schielzeth, 2011). P-values for the individual effects were based on likelihood ratio tests comparing the full model with respective reduced models (R function drop1). Performance within a given condition was compared against chance (50% correct) by fitting a GLMM with centered predictors and testing whether the intercept differed from zero. All models were fitted in R (R Core Team, 2016) using the function lmer of the R-package lme4 (Bates et al., 2015). The code along with the raw data can be found in the online materials. On an individual level, performance in each condition was considered above chance if 10 or more trials (out of 12) were correct (binomial test, $p < .05$).

Results

Overall, the full model was significant compared to the null model ($X^2(31, N=672) = 15$, $p = .003$). To simplify the model, all interactions and their random slopes were removed, as they were found to be non-significant (see Table A2). The final model included only the main effects of condition, session number, trial number, age and age^2.

Subjects´ choice was significantly influenced by the test condition (GLMM; $X^2(3, N=672) = 27.78$, $p < .001$). In other words, chimpanzees chose differently depending on the content of the two apparatuses´ containers. There was no effect of session number ($X^2(1, N=672) = 0.810$, $p = .37$) or trial number ($X^2(1, N=672) = 0.10$, $p = .75$), suggesting that chimpanzees' performance did not change with increasing experience within a session or over the course of sessions (see Table A3 for more details).

Proportional Information

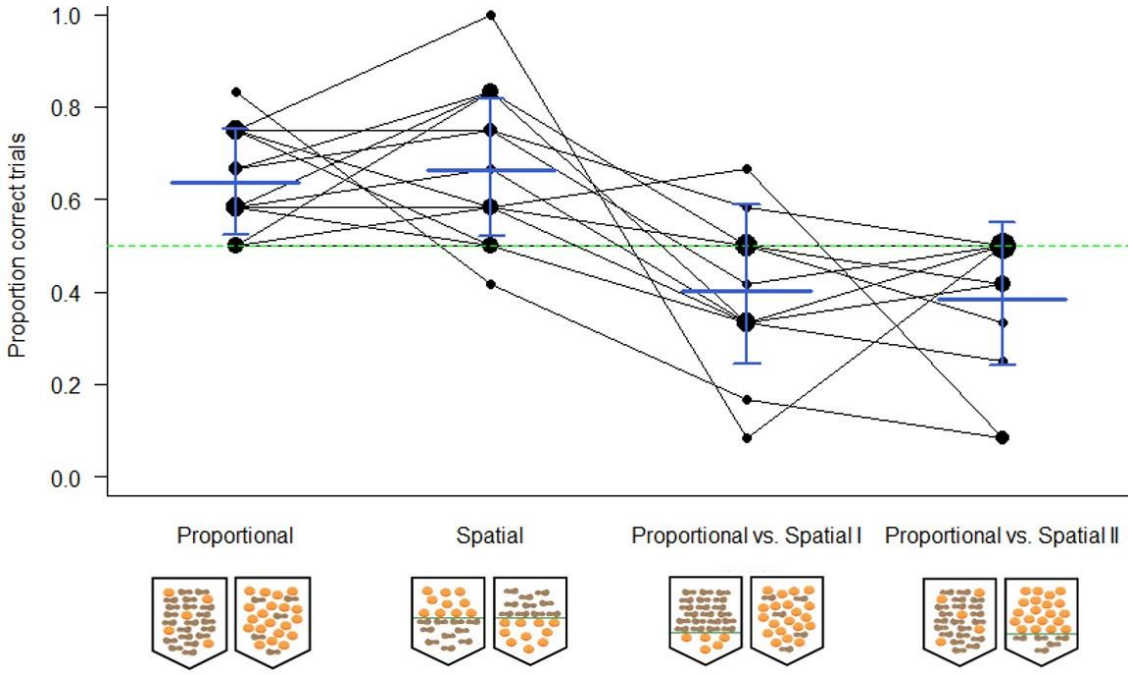
When the food items within both populations were randomly mixed, chimpanzees as a group chose to operate the apparatus containing the greater proportion of peanuts to carrots more often than what would have been expected by chance ($Mean_{favorable\ apparatus}$ = 64.3% of trials; $\beta = 0.60$; $p < .001$; 95% CI [0.123, 1.077]; see Figure 3). On the individual level, eleven subjects chose correctly in more than 50% of trials, one of them significantly above chance level. No individual performed below chance level (see Table A1). These results replicate previous findings in demonstrating that most chimpanzees are able to draw inferences from populations of food items to randomly drawn samples (Eckert, Call et al., 2018; Eckert, Rakoczy et al., 2018; Rakoczy et al., 2014). Additionally, the results suggest that this capacity is independent of the task setup: in contrast to all previous studies, chimpanzees in our study acted as sampling agents and actively drew the samples themselves, rather than picking a sample drawn by the experimenter.

Spatial Information

When both populations contained the same proportions of preferred to non-preferred food items, chimpanzees as a group chose to operate the apparatus in which the preferred food items were accessible (i.e., placed underneath, rather than above the physical barrier) more often than what would have been expected by chance ($Mean_{favorable\ apparatus}$ = 66.7% of trials; $\beta = 0.73$; $p < .001$; 95% CI [-0.550, 0.801]; see Figure 3). On the individual level, ten subjects chose correctly in more than 50% of trials, four of them performed significantly above chance level. Again, no individual performed significantly below chance level (see Table A1). These results show that most chimpanzees understood the workings of the apparatus and were aware of the effect that the spatial distribution had on the outcome of the sampling event. In other words, chimpanzees seemed to understand that food items close to the exit, (i.e., below the barrier), were the ones that would be sampled.

Figure 3

Chimpanzees´ Performance Across Conditions



Note. Shown is the proportion of trials in which subjects chose the sample from the apparatus that was more likely to deliver a preferred food item as a sample. The dot size indicates the number of subjects performing at the same level. The dashed green horizontal line depicts chance level. Bold blue horizontal lines depict the mean probability predicted by the model and blue vertical lines depict bootstrapped 95% confidence intervals.

Proportional vs. Spatial Information I

When both populations presented a low chance of delivering a preferred food item as a sample, chimpanzees as a group chose the apparatus containing the greater proportion of preferred to non-preferred food items, despite the fact that the preferred food items in this apparatus were inaccessible (and thus chances of receiving a preferred food item were 0%) ($Mean_{favorable\ apparatus}$ = 41.07% of trials; β = -0.37; p = .024; 95% CI [-1.733, -0.246]; see Figure 3). On the individual level, two subjects chose correctly in more than 50% of trials. None of those performed significantly above chance level. Two subjects performed significantly below chance level (see Table A1). In sum, we did not find any evidence that chimpanzees correctly integrated proportional and physical information in this condition.

Proportional vs. Spatial Information II

When both populations yielded a high chance of delivering a preferred food item as a sample, chimpanzees as a group choose the apparatus containing the greater proportion of preferred to non-preferred food items, despite the fact that objective chances for a preferred food item as a sample were only 80% in that apparatus, whereas they were 100% in the other apparatus ($Mean_{favorable\ apparatus}$= 39.29% of trials; β = -0.44; p = .007; 95% CI [-1.676, -0.491]; see Figure 3). On the individual level, no subject chose correctly in more than 50% of trials. Two individuals performed significantly below chance level (see Table A1). Thus, chimpanzees did not seem to integrate proportional and physical information in this condition.

Discussion

We found that, as a group, chimpanzees in the current task used proportional information to infer which population was more likely to lead to a preferred food item as a randomly drawn sample. Furthermore, most chimpanzees also used spatial information to judge the outcome of a sampling event. However, when proportional and spatial information were in conflict, chimpanzees failed to correctly integrate the two sources of knowledge. In these circumstances, they based their choice on the proportional composition of the population, disregarding the spatial distribution.

These findings seem to contradict, firstly, findings in analogous tasks with human infants (Denison et al., 2014; Denison & Xu, 2010b; Lawson & Rakison, 2013; Teglas et al., 2007, 2011) and kea (Bastos & Taylor, 2020), and secondly, chimpanzees' performance in conceptually similar tasks in which statistical and *social* information needed to be integrated. When presented with biased experimenters, chimpanzees did correctly combine knowledge about others' behavior and mental states with their intuition of statistical regularities to draw rational inferences about the outcome of a sampling event (Eckert, Rakoczy et al., 2018). This suggested that great apes, like human infants and kea, might possess a domain-general statistical inference mechanism. However, a characteristic of domain-generality is that different types of information from different cognitive domains, both social and physical, can be utilized and combined to make rational judgements. Does the failure of chimpanzees to reason about physical information when drawing statistical inferences thus imply that apes are not capable of domain-general statistical inference? Or do these negative findings reflect mere performance limitations, rather than competence limits? We will discuss five possible explanations (in terms of performance limitations based on task demands) that might account for the apes' poor performance in the crucial test conditions, in which proportional information and physical information were conflicted.

A first possibility is that the current task setup, in contrast to the previously used "bucket paradigm" (Eckert, Call et al., 2018; Eckert, Rakoczy et al., 2018; Rakoczy et al., 2014) was not sufficiently intuitive for the chimpanzees. Perhaps they would need more experience with the workings of the gumball machines in order to fully understand the influence of the spatial distribution on the sampling outcome. However, while it is true that, unlike previous studies, our task required a basic understanding of naïve physics, especially the laws of gravity, it seems unlikely that this was the limiting factor in the current study for several reasons. First, the apes' success in both the "proportional information" and the "spatial information" condition shows that chimpanzees did understand the apparatus to some degree, and that it was the integration of spatial and proportional information which induced poor performance. Second, comparative work has repeatedly demonstrated that chimpanzees and other apes excel in tasks involving basic physical reasoning (e.g., Herrmann et al., 2007; Seed & Tomasello, 2010; Tomasello & Call, 1997; Wobber et al., 2014), including those requiring a basic understanding of gravity (Cacchione et al., 2009; Cacchione & Call, 2010; Cacchione & Krist, 2004). Third, prior to the experiment reported here, chimpanzees had already participated in a similar version of the same task, using the same apparatuses (see Supplementary Material). In this previous version, chimpanzees performed poorly, which is why we decided to test them again with a modified (i.e., simplified) test procedure. Lastly, our analysis revealed that chimpanzees neither became more proficient within one test session (i.e., within a condition), nor over the course of several test sessions. Therefore, it seems unlikely that chimpanzees´ poor performance can be explained by a lack of experience with the apparatuses or the procedure.

A second possible reason for chimpanzees failing to acknowledge the physical determinants in the integration conditions is that choosing the proportionally favorable population was always rewarded, both in the "Proportional information" condition of the current study, as well as in a previous study on intuitive statistics with the same subjects (Eckert, Call et al., 2018). It is thus possible that chimpanzees generalized this knowledge to the integration conditions, and, therefore, ignored the physical constraints determining the outcome in these conditions. However, we believe that such generalization across conditions and studies is highly unlikely for three reasons. First, in the current study, subjects´ performance did not change with experience within the "Proportional information" condition. If chimpanzees had learned through the rewarding pattern that choosing the proportionally favorable

population was always correct, their performance should have increased across trials. Such a performance increase was not found in our study. Second, in none of the previous studies on intuitive statistics in chimpanzees (Eckert et al., 2017; Eckert, Call et al., 2018; Eckert, Rakoczy et al., 2018; Rakoczy et al., 2014) we found any evidence for learning or association effects. Instead, in all studies to date, chimpanzees' choice in the very first trial of a condition was representative of their overall performance. It is thus highly unlikely that subjects formed such an association in the current study. Lastly, in one of these previous studies (Eckert, Rakoczy et al., 2018), the same individuals as in the present study did intuitively disregard proportional information in favor of information about the experimenters' behavior and mental states when appropriate, even though they had been rewarded for choosing according to proportions in a previous study (Eckert, Call et al., 2018). Chimpanzees' decisions in that study were modulated by very fine-grained differences in the experimenters' behavior (whether they looked into the buckets while sampling or drew blindly), suggesting that chimpanzees did pay close attention to the scene, rather than simply picking what had been repeatedly rewarded before. Therefore, while it is certainly not a trivial task to overcome a previously rewarded choice, we think that this difficulty alone does not account for the chimpanzees' failure in the current study.

A third potential explanation for poor performance in the integration conditions is that chimpanzees may have been distracted by the large amount of preferred food items in the distractor apparatus. More specifically, in these conditions, the "correct" apparatus contained the smaller number of peanuts compared to the "incorrect" apparatus (but due to the barrier they were inaccessible for sampling). It is well known from previous research that chimpanzees and other non-human primates have great difficulties in overcoming their natural tendency to reach for the larger of two amounts of food (Beran et al., 2016; e.g., Boysen & Berntson, 1995; Schmitt & Fischer, 2011; Vlamings et al., 2006). In our study, chimpanzees were successful in only two of the four test conditions, namely in those in which it was not necessary to inhibit the "go for more" tendency: in the "proportional information" condition, the correct answer was to choose the apparatus containing more peanuts. In the "spatial information" condition, there could be no interference because both apparatuses contained the same numbers of peanuts. By contrast, in the two integration conditions, apes were required to resist operating the apparatus containing the larger number of peanuts. Hence, it is plausible that in our study, chimpanzees' natural tendency to choose the apparatus containing more peanuts interfered with their ability to make correct inferences regarding the most likely outcome of a sampling event. However, in a previous study on intuitive statistics, the same group of chimpanzees did not show signs of interference when the "correct" answer was to choose the sample coming from the population with absolutely fewer preferred food items (Eckert, Rakoczy et al., 2018). In that study, chimpanzees correctly chose a sample coming from a population with fewer preferred food items, when the sampling agent had previously expressed a preference for preferred food items (and had visual access to the population while sampling) and was therefore likely to sample a preferred food item again. Hence, it seems that under some circumstances chimpanzees are able to overcome their natural tendency to pick the larger of two amounts of food in tasks requiring statistical inference.

A fourth possible explanation for the discrepancy between chimpanzees' performance in the current study and performance in previous, conceptually similar studies is our task setup. Whereas in previous studies subjects were passively watching the sampling event and could choose between the two samples once the experimenter had drawn them, in the current setup chimpanzees were actively involved in the procedure and (seemingly) drew the samples themselves. It is conceivable that this active role in picking one of the populations, instead of one of the already drawn samples, facilitated their tendency to "go for more," and hence to overlook the spatial composition of the food items in the crucial conflict conditions. In the recent study showing that kea can consider spatial information when drawing statistical inferences (Bastos & Taylor, 2020), birds were not required to draw samples themselves, but chose between two samples that had been drawn by an experimenter. Similarly, in developmental research demonstrating human infants' competence in integrating statistical and physical information (Denison et al., 2014; Denison & Xu, 2010b; Teglas et al., 2007, 2011), infants never acted as sampling agents themselves. Future research should therefore investigate whether chimpanzees would successfully

integrate spatial and proportional information in a similar task setup that does not involve an active sampling role for the subjects.

A fifth possible explanation for chimpanzees' failure to correctly integrate proportional and spatial information is that the difference in probabilities for receiving a preferred food item between the two populations was too small in the crucial conditions. In the "Proportional vs. spatial information I" condition, the probability of sampling a peanut was 0% in one population, and 20% in the other (i.e., an absolute difference of 20%). Similarly, in the "Proportional vs. spatial information II" condition, the probability of sampling a peanut was 100% in one population, and 80% in the other (also an absolute difference of 20%). It is possible that this difference was too subtle to be detected by the chimpanzees. Previous research found that performance in probabilistic reasoning tasks involving a choice between two variable options is highly dependent on the difference between probabilities (Eckert, Call et al., 2018; Hanus & Call, 2014). In one study (Eckert, Call et al., 2018) chimpanzees successfully chose a sample coming from the proportion-wise more favorable population only when the absolute difference in likelihood for obtaining the preferred sample between both populations was 33% or larger. When the likelihoods differed by only 17% or less, chimpanzees performed at chance level. Similarly, Hanus and Call (2014) found that chimpanzees' performance in a probabilistic reasoning task was directly influenced by the relative difference between two given probabilities (i.e., by the probability ratio) and that chimpanzees only preferred the more likely option once a certain threshold (a probability ratio of 0.33) was reached. Considering these previous findings, it is conceivable that the probability ratio in the integration conditions in our study was too subtle and may have been well below chimpanzees' threshold for probability discrimination.

Chimpanzees' failure in the integration conditions reveals performance limitations introduced by our task setup and suggests that the apes might lack an important feature of logical thought: the ability to recognize the special value of a truly safe (or truly bad) option, and the implication that one can ignore other types of information once one knows the outcome of one of two options with certainty. Instead, the apes in our study still seemed to rely on inferences produced by probabilistic reasoning based on proportions when the spatial arrangement of food items in one option created certainty as to which outcome will happen. This finding is in line with what has been found in the above-mentioned study on probabilistic reasoning (Hanus & Call, 2014); similarly to our integration conditions, chimpanzees in this study did not appreciate the utility of a truly safe option (i.e., an option with 100% chance of obtaining the reward). Instead, they systematically preferred the more likely option only once the threshold probability ratio was reached. These results are in contrast to findings on human adults (Tversky & Kahneman, 1979; Weller et al., 2011), who have a strong preference for options that offer a safe reward compared with any other option, provided the maximum outcome is the same. Hanus and Call (2014) interpreted their findings as chimpanzees being guided by intuitive mathematics ("How much do the two likelihoods differ") rather than by intuitive logic ("Is one option safe?"). Similar findings have also been obtained in research with young children (Mody & Carey, 2016; also see Leahy & Carey, 2020, for a review), suggesting that this type of logical inference might emerge only later in human ontogeny, and may not be shared with our closest living relatives.

Looking at differences between individuals´ performances, we found a relatively coherent pattern in line with our statistical analysis: In the Proportional condition, all individuals chose correctly in at least 50% of trials. Similarly, in the Spatial condition, only one individual chose correctly in less than 50% of trials. In the two integration conditions, by contrast, no individual (Proportional vs. Spatial II) and no but two individuals (Proportional vs. Spatial I) chose correctly in more than 50% of trials. The (rather subtle) differences between individuals may in part be situational: Each condition was tested on one day only. If a subject was distracted on that particular day (e.g., because of social events in the group) they may have performed worse than they would have on an average day, and hence their performance in the respective condition may have suffered. It is also possible, however, that the differences between individuals are stable. For example, individual differences in temperament or differences in food motivation could drastically impact subjects' performances. Within the current study, we cannot discriminate between situational and permanent individual differences. It would be of great interest for future research to

compare individual performances across related studies to explore whether certain performance patterns are consistent across time and tasks.

Although the findings of the current study may suggest that chimpanzees have difficulties correctly integrating statistical and physical information, future research should also investigate another aspect of domain-generality: context and stimulus independence. Human infants not only integrate different types of information into their statistical inference, but they also compute probabilities in a variety of different contexts using various types of stimuli (e.g., candy, toys, and arbitrary objects; Denison & Xu, 2010b, 2014; Gweon et al., 2010; Gweon & Schulz, 2011; Teglas et al., 2011). In comparative research, by contrast, all studies testing statistical reasoning abilities in great apes used food items as stimuli over which probabilities had to be computed (Eckert et al., 2017; Eckert, Call et al., 2018; Eckert, Rakoczy et al., 2018; Rakoczy et al., 2014). If chimpanzees, like human infants, possessed a domain-general statistical reasoning mechanism, we would expect their abilities to extend to computing probabilities over other types of stimuli (e.g., frequencies of events, such as successes vs. failures). Therefore, while in the current study we focused on one aspect of domain-generality (i.e., cross-domain information integration), it will be an interesting question for future research to explore whether apes' statistical reasoning abilities show signatures of another aspect of domain-generality (i.e., context and stimulus independence). However, it is likely that at least chimpanzees will also display context and stimulus independence given their ability to process quantitative information about food items, objects, stimuli on a computer screen and even the number of individuals present in a simulated territorial encounter (e.g., Beran, 2006; Hanus & Call, 2007; Matsuzawa, 1985; Wilson et al., 2012).

In sum, we believe that the observed performance patterns do, to some extent, reflect performance limitations, such as an inability to detect the probability difference in combination with a natural tendency to choose the larger of two amounts of (preferred) food, which may have been enhanced by the fact that chimpanzees extracted the samples themselves. At the same time, apes' performance patterns also support the hypothesis that chimpanzees have some striking limitations in their cognitive competency regarding logical conclusions: as in a previous study (Hanus & Call, 2014) chimpanzees failed to consider certainty in absolute terms (in our case created by spatial arrangement) when it conflicted with proportional information. It will be an interesting question for future research to investigate whether the prevalence of intuitive statistics based on proportional information over logic has limits, and if chimpanzees would, given a large enough probability difference, abandon statistical inference in favor of logic.

Conclusion

In this study we used a novel task to investigate chimpanzees' capacity to reason from population to sample. Chimpanzees acted as sampling agents operating a gumball machine by themselves. We found that in this setup subjects chose to operate the apparatus that was more likely to deliver a preferred food item as a sample both when the proportion of preferred to non-preferred food items determined the outcome, and when the spatial composition of food items within an apparatus determined the outcome. When, however, proportional and spatial composition of populations were in conflict, chimpanzees failed to correctly integrate the two sources of knowledge, and based decisions on proportional composition alone. We suggest that these findings reflect performance limitations (the inability to detect the subtle probability differences between the two options in the critical conditions, as well as challenges of inhibitory control with this task setup) and competence limitations (the failure to recognize the special value of a truly safe option when it conflicts with statistical inference based on proportions). Future research will need to address these limitations in order to shed more light on a possible domain-general statistical reasoning mechanism in great apes.

Acknowledgements

Funding: this work was supported by a research grant of the German Science Foundation DFG (grant # RA 2155/3-1) to Hannes Rakoczy and Josep Call. We acknowledge additional support by the Leibniz Association through funding for the Leibniz ScienceCampus Primate Cognition. We are thankful to Chimpanzee Sanctuary and Wildlife Conservation Trust and especially all keepers involved in our study for providing us with the opportunity to test at Ngamba Island. We also appreciate permission from the Ugandan National Council for Science and Technology and the Uganda Wildlife Authority. We thank the MPI EVA Multimedia Department, in particular Franziska Honigschnabel, for preparing the drawings of the task setup. We are grateful to Manuel Bohn for statistical advice. Lastly, we thank the two anonymous reviewers for valuable feedback which helped us to tremendously improve the manuscript.

Data Availability

Accompanying data and R-code are available online.

Competing Interests

The authors declare that they have no competing interests.

References

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

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Bastos, A. P. M., & Taylor, A. H. (2020). Kea show three signatures of domain-general statistical inference. *Nature Communications*, *11*(1), 1–8. https://doi.org/10.1038/s41467-020-14695-1

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Beran, M. J. (2006). Quantity perception by adult humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*), and rhesus macaques (*Macaca mulatta*) as a function of stimulus organization. *International Journal of Comparative Psychology*, *19*(4), 386–397.

Beran, M. J., James, B. T., Whitham, W. W., & Parrish, A. E. (2016). Chimpanzees can point to smaller amounts of food to accumulate larger amounts but they still fail the reverse-reward contingency task. *Journal of Experimental Psychology: Animal Learning and Cognition*, *42*(2), 347–358. https://doi.org/10.1037/xan0000115

Boysen, S. T., & Berntson, G. G. (1995). Responses to quantity—Perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology-Animal Behavior Processes*, *21*(1), 82–86. https://doi.org/10.1037/0097-7403.21.1.82

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences*, *40*, e195. https://doi.org/10.1017/S0140525X16000959

Cacchione, T., & Call, J. (2010). Intuitions about gravity and solidity in great apes: The tubes task. *Developmental Science*, *13*(2), 320–330. https://doi.org/10.1111/j.1467-7687.2009.00881.x

Cacchione, T., Call, J., & Zingg, R. (2009). Gravity and solidity in four great ape species (*Gorilla gorilla*, *Pongo pygmaeus*, *Pan troglodytes*, *Pan paniscus*): Vertical and horizontal variations of the table task. *Journal of Comparative Psychology*, *123*(2), 168. https://doi.org/10.1037/a0013580

Cacchione, T., & Krist, H. (2004). Recognizing impossible object relations: Intuitions about support in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *118*(2), 140–148. https://doi.org/10.1037/0735-7036.118.2.140

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(6), 657–674. https://doi.org/10.1017/S0140525X02000122

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

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

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

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

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

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

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

Denison, S., & Xu, F. (2010b). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, *13*(5), 798–803. https://doi.org/10.1111/j.1467-7687.2009.00943.x

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. *Cognition*, *130*(3), 335–347. https://doi.org/10.1016/j.cognition.2013.12.001

Dobson, A. J., & Barnett, A. G. (2008). *An introduction to generalized linear models, third edition*. CRC Press.

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

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

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

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE.

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology & Sociobiology*, *65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. SAGE Publications.

Gentner, D., Holyoak, K. J., Holyoak, K. J., & Kokinov, B. N. (2001). *The Analogical Mind: Perspectives from Cognitive Science*. MIT Press.

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

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

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

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, 201003095. https://doi.org/10.1073/pnas.1003095107

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

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

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial orientation. *Cognition*, *61*(3), 195–232. https://doi.org/10.1016/S0010-0277(96)00714-7

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*(5843), 1360–1366. https://doi.org/10.1126/science.1146282

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. In C. A. S Staël Von Holstein (Ed.), *The concept of probability in psychological experiments. Theory and decision library* (Vol 8). Springer. https://doi.org/10.1007/978-94-010-2288-0_3

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

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

Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences*, *24*(1), 65–78. https://doi.org/10.1016/j.tics.2019.11.004

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

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

Mithen, S. (1996). The prehistory of the mind: The cognitive origins of art and science. *Thames & Hudson Ltd.*

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40–48. https://doi.org/10.1016/j.cognition.2016.05.012

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Prentice-Hall.

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

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), 181025. https://doi.org/10.1098/rsos.181025

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*(1), 1–12. https://doi.org/10.1038/s41598-019-48543-0

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

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

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

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

Seed, A., & Tomasello, M. (2010). Primate cognition. *Topics in Cognitive Science*, *2*(3), 407–419. https://doi.org/10.1111/j.1756-8765.2010.01099.x

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

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

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

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

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

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

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

Tversky, A., & Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–291.

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

Vlamings, P. H. J. M., Uher, J., & Call, J. (2006). How the great apes (*Pan troglodytes, Pongo pygmaeus, Pan paniscus, and Gorilla gorilla*) perform on the reversed contingency task: The effects of food quantity and food visibility. *Journal of Experimental Psychology-Animal Behavior Processes*, *32*(1), 60–70. https://doi.org/10.1037/0097-7403.32.1.60

Weller, J. A., Levin, I. P., & Denburg, N. L. (2011). Trajectory of risky decision making for potential gains and losses from ages 5 to 85. *Journal of Behavioral Decision Making*, *24*(4), 331–344. https://doi.org/10.1002/bdm.690

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

Wilson, M. L., Kahlenberg, S. M., Wells, M., & Wrangham, R. W. (2012). Ecological and social factors affect the occurrence and outcomes of intergroup encounters in chimpanzees. *Animal Behaviour*, *83*(1), 277–291. https://doi.org/10.1016/j.anbehav.2011.11.004

Wobber, V., Herrmann, E., Hare, B., Wrangham, R., & Tomasello, M. (2014). Differences in the early cognitive development of children and great apes: Cognitive development in humans and great apes. *Developmental Psychobiology*, *56*(3), 547–573. https://doi.org/10.1002/dev.21125

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

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

Appendices

Appendix Video 1a and 1b Illustration of the workings of the apparatus

https://share.eva.mpg.de/index.php/s/LtNGyStTBeEzpe7
https://share.eva.mpg.de/index.php/s/ZRamcyz8qCZ5reD

Appendix Video 2 Example trial of the test condition "Spatial information"

https://share.eva.mpg.de/index.php/s/wNP2HDQifDkimYn

Table A1

Detailed Subject Information

Subject name	Sex	Age	Participation in previous studies on intuitive statistics		Order of Conditions*	Performance [Proportion of correct trials]			
			Eckert, Call, et al., 2018	Eckert, Rakoczy, et al., 2018		Proportional	Spatial	Proportional vs. Spatial I	Proportional vs. Spatial II
Asega	m	19	yes	yes	3,2,1,4	0.50	0.58	0.50	0.50
Baluku	m	19	yes	yes	1,3,2,4	0.50	0.83	0.33	0.42
Becky	f	27	yes	yes	3,4,1,2	0.58	0.67	0.33	0.50
Bili	f	19	yes	no	4,3,2,1	0.58	0.50	0.50	0.50
Bwambale	m	17	yes	yes	2,4,1,3	0.58	0.58	0.67	0.08
Cocoa	f	10	yes	yes	1,4,3,2	0.75	0.75	0.42	0.50
Kidogo	f	33	yes	yes	3,2,4,1	0.75	0.50	0.33	0.25
Mawa	m	21	yes	yes	1,2,3,4	0.50	0.50	0.50	0.42
Medina	f	10	yes	yes	2,3,4,1	0.67	0.83	0.50	0.33
Nakuu	f	16	yes	yes	4,1,3,2	0.58	0.83	0.50	0.50
Nani	f	16	yes	yes	2,4,3,1	0.75	0.58	0.33	0.42
Nkuumwa	f	21	yes	yes	2,4,3,1	0.75	1.00	0.08	0.50
Pasa	f	18	yes	yes	1,2,4,3	0.67	0.75	0.58	0.50
Yoyo	f	19	yes	yes	2,1,4,3	0.83	0.42	0.17	0.08

*1=Proportional; 2=Spatial; 3=Proportional vs. Spatial I; 4=Proportional vs. Spatial II

Table A2

Effect of Terms Removed from the Model

Term	X^2	df	*p*
condition*session number* trial number	2.230	3	.526
condition*session number	0.591	3	.898
condition*trial number	3.109	1	.375
session number*trial number	0.043	1	.835

Note. The full model was significant compared to the null model (X^2 = 34.156, df=15, p < .01). The interactions between condition, session number and trial number were found to have no significant effect on subjects´ choice and were therefore removed from the final model.

Table A3

Influence of Terms Included in the Final Model on Subjects´ Choice

Term	Estimate	SE	X^2	Df	*p*
intercept	0.56	0.17	27.78	3	.001
condition (spatial)[(1)]	0.12	0.25	27.78	3	.623
condition (proportional vs. spatial I)[(1)]	-0.97	0.26	27.78	3	<.001
condition (proportional vs. spatial II)[(1)]	-1.03	0.23	27.78	3	<.001
session[(2)]	-0.03	0.09	0.10	1	.754
trial[(2)]	0.07	0.08	0.81	1	.368
age[(2)]	-0.15	0.09	2.44	1	.118
$(age)^{2}$[(2)]	0.03	0.06	0.26	1	.613

Note. (1) the condition “proportional” served as reference category; (2) these predictors were z-transformed