Title: Intuitive statistical inferences in chimpanzees and humans follow Weber's Law

Johanna Eckert[1,2,3], Josep Call[1,4], Jonas Hermes[2,3], Esther Herrmann[1], Hannes Rakoczy[2,3]

[1] Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

[2] Department of Developmental Psychology, University of Goettingen, Waldweg 26, 37073 Goettingen, Germany

[3] Leibniz ScienceCampus "Primate Cognition", German Primate Center/Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Goettingen, Germany

[4] School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, UK

Corresponding Author: Johanna Eckert

Email: johanna.eckert@eva.mpg.de

Postal address: Deutscher Platz 6, 04103 Leipzig, Germany

Phone +49 (341) 3550 – 424

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

Abstract

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

1. Introduction

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

59   objects´ spatial distance from the opening (the closer objects are, the more likely they are to exit;

60   Teglas et al., 2011).

61   These findings demonstrated that neither language nor mathematical education are

62   prerequisites for basic statistical reasoning. Instead, infants seem to be equipped with a set of

63   efficient and powerful statistical inference mechanisms from very early on, which presumably help

64   them to rapidly learn about rules and regularities of the world. From a comparative point of view,

65   these findings raise the question of whether intuitive statistics may be part of our evolutionary

66   heritage and therefore not necessarily uniquely human.

67   To shed light on this question, comparative research adapted Denison and Xu´s (2010)

68   paradigm and tested nonhuman great apes for their statistical reasoning capacities: Rakoczy et al.

69   (2014) presented chimpanzees, gorillas, orangutans and bonobos with two transparent buckets

70   containing mixtures of preferred and non-preferred food items (banana pellets and carrot pieces)

71   in specific ratios. Subsequently, the experimenter drew one sample each and the subject was

72   allowed to choose between the two covered samples. Apes were able to infer which of the two

73   populations was more likely to lead to a pellet as a sample. Moreover, they chose systematically

74   even when absolute and relative frequencies were disentangled, i.e. when the population with the

75   more favorable ratio of pellets to carrots contained absolutely fewer pellets than the other one. A

76   very recent study (Eckert, Rakoczy, Call, Herrmann, & Hanus, 2018) showed that these statistical

77   inferences are not an isolated and automatic process; instead, just like human infants, apes did

78   consider additional information about the experimenter (her preferences and visual access) when

79   predicting the outcome of her draw. Hence, some great ape intuitive statistical abilities seem to be

80   on a par with those of human infants, suggesting that they constitute an evolutionary ancient

81   ability.

82   However, when great apes were presented with the reverse task, requiring inferences from

83   sample to population, they exhibited some limitations: Eckert, Rakoczy, and Call (2017) presented

84   apes with covered containers holding populations of preferred and non-preferred food-items. After

85   observing multi-item samples being drawn from these populations, apes could choose between the

86   two covered containers. Subjects were able to correctly reason from sample to population, but only

87   in conditions, in which the proportionally favorable sample also contained absolutely more

88   preferred food items than the other (4:1 vs. 1:4 preferred to non-preferred food items). In

89   experiments contrasting absolute and relative frequencies of preferred food items (e.g. 2:1 vs. 4:8

90   preferred to non-preferred food items), apes tended to choose the population from which the

91   sample with absolutely more preferred items was drawn (4:8), despite its unfavorable ratio.

92   There are at least two possible interpretations of these findings: First, they may indicate

93   that apes relied on absolute quantity heuristics to reason from sample to population. In fact, even

94   in the human literature there is a great deal of research suggesting that most probabilistic inferences

95   are actually just the result of different heuristics (e.g. Davidson, 1995; De Neys & Vanderputte,

96   2011; Jacobs & Potenza, 1991; Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1974,

97   1981). Hence, controlling for absolute number heuristics is of substantial importance in the field

98   of intuitive statistics. Nevertheless, many studies (both on human and nonhuman species) have

99   unfortunately failed to sufficiently control for such strategies. The previously described studies on

100   great apes, for instance, included control conditions for heuristics dealing with the absolute

101   quantity of preferred items (which apes passed in Rakoczy et al. (2014), but failed in Eckert et al.

102   (2017)). None of them, however, tested for the reverse strategy: a heuristic based on avoiding the

103   population or sample containing more non-preferred food items. Hence, to be able to draw

104　conclusions about apes´ intuitive statistical abilities, there is an urgent need for studies controlling

105　for all types of absolute quantity heuristics.

106　　　　A second interpretation for Eckert et al.´s (2017) negative findings is that they merely

107　reflect performance, rather than competence limitations. The critical conditions in that study may

108　have been especially difficult because the magnitude of difference between samples (i.e. the ratio

109　of the two ratios, ROR) was relatively small, and perhaps beyond the signature limits of apes´

110　capacity: While the ROR was 16 in the successful confounded conditions (and also in all

111　conditions of Rakoczy and colleagues´ population to sample study), it was as low as 4 in the critical

112　condition in which absolute and relative frequencies of preferred food items were disentangled.

113　　This, in turn, raises a much more fundamental question: What are the cognitive foundations of

114　intuitive statistics that explain both the scope and limits of this capacity? And are these the same

115　in humans and our closest living relatives, the chimpanzees?

116　　　　Nonhuman primates (and other animals) share with humans a cognitive mechanism for

117　basic quantitative cognition. This mechanism, the analogue magnitude system, is used for dealing

118　with absolute numerical information: It represents number (and also other magnitudes like duration

119　or space) by a mental magnitude that is roughly proportional to and thus a direct *analogue* of the

120　number of individuals in the set being enumerated (see, e.g. Carey, 2009; Dehaene, 2011).[1] It thus

121　enables subjects to estimate and compare arbitrarily large quantities, but only in an approximate

122　way (e.g. Nieder & Dehaene, 2009). Its accuracy follows Weber´s Law: Discriminability of two

---

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

123   sets varies as a function of the ratio of the set sizes to be compared, independently of their absolute

124   numerosity (e.g. Cantlon & Brannon, 2006, 2007). For example, if a subject can discriminate 2

125   from 4 objects, it is also able to discriminate 10 from 20 or 500 from 1000. This fundamental

126   characteristic yields specific signatures that can be used to identify the involvement of this system

127   in cognitive tasks. Numerous comparative studies have shown that many species across the animal

128   kingdom exhibit the same signatures in accordance with Weber´s law when confronted with

129   quantity comparison tasks. For example, fish (Buckingham, Wong, & Rosenthal, 2007), birds

130   (Ain, Giret, Grand, Kreutzer, & Bovet, 2009; Rugani, Cavazzana, Vallortigara, & Regolin, 2013),

131   monkeys (Barnard et al., 2013; Cantlon & Brannon, 2007), and great apes (Beran, 2004; Call,

132   2000; Hanus & Call, 2007) all showed a ratio-dependent performance when discriminating

133   absolute quantities, suggesting that the analogue magnitude system is an evolutionary ancient

134   mechanism (see Beran, 2017 for a review).

135   Are the limits reported by Eckert et al. (2017) a first hint that apes employed their analogue

136   magnitude system in a statistical reasoning task as well? Is the ability to reason probabilistically

137   from population to sample and vice versa dependent on the ratio between the two proportions to

138   be discriminated, i.e. on the ratio of ratios (ROR)? To our knowledge, no study has directly tested

139   for this hypothesis in nonhuman primates yet; indeed, even in human adults the evidence is

140   ambiguous. One study (O´Grady, Griffiths, & Xu, 2016) tested human adults in a computer based

141   task that required them to reason statistically from population to sample. The authors varied the

142   magnitude of difference between the two populations´ ratios and included both trials in which

143   absolute and relative frequencies of target stimuli were confounded and trials in which they were

144   disentangled. Statistical analysis revealed that the effect of the ROR on humans´ performance was

145   dependent on the trial type, with much stronger effects in the confounded condition. In this

146   condition, participants could simply compare the two (absolute) amounts of target stimuli, a

147   capacity known to be enabled by the analogue magnitude system. Unfortunately, it remained

148   unclear whether there was any significant ROR impact in the crucial trials controlling for absolute

149   quantity heuristics.[2] Hence, this study is yet another example of the problematic confound of

150   absolute and relative frequencies in intuitive statistical reasoning tasks.

151   The only two studies investigating signatures in a statistical reasoning task in human infants

152   used looking-time patterns and produced mixed results. On the one hand, Téglás, Ibanez-Lillo,

153   Costa and Bonatti (2015) found that, after watching a scene containing moving objects of two

154   ensembles, infants looked longer at an unlikely than at a likely single-case outcome when the scene

155   depicted a 3:1 ratio. However, they did not do so with a 12:4 ratio, suggesting that absolute set

156   sizes, rather than ratios, influenced performance in this intuitive statistical reasoning task. On the

157   other hand, Kayhan, Gredebäck and Lindskog (2017) measured infants´ looking patterns at two

158   multi-item-samples drawn from one population. In order to vary the magnitude of difference in

159   likelihoods between samples, they manipulated both the ratios within samples and within the

160   population. Here, results showed that infants´ looking patterns varied as a function of the

161   magnitude of difference in likelihood, suggesting that ROR does modulate infants´ probability

162   estimations. Hence, to date it remains unclear whether or not the analogue magnitude system is

163   the primary cognitive mechanism enabling intuitive statistics in humans.

---

[2] In fact, the authors found a significant three-way-interaction between ROR, condition and age. Hence, the effect of the ROR was not only dependent on the trial type, but also on the age of participants (yet this age effect was not discussed in the paper). It is generally not meaningful or reasonable to interpret the individual effects of the components of a significant interaction (Bortz, 1999; Underwood, 1997; Zar, 1999). The only legitimate statement this study can draw regarding ROR effects is, therefore, that the effect was only visible in certain age groups depending on whether or not absolute and relative frequencies were confounded.

164   Despite this controversy, there is some indirect evidence supporting the idea of the

165   analogue magnitude system as foundation of statistical reasoning: Both developmental and

166   comparative research have shown that one important prerequisite capacity for statistical reasoning

167   – tracking relative frequencies – is subject to the same signatures as tracking absolute frequencies.

168   McCrink and Wynn (2007) presented human infants with a ratio discrimination task: After

169   habituating them with multiple examples of a single ratio, infants were able to discriminate

170   between new examples of this ratio and novel ratios. Infants´ accuracy was highly dependent on

171   the ratio between ratios, in accordance with Weber´s law. Similar results were found for a

172   nonhuman primate species (Drucker, Rossa, & Brannon, 2015): two rhesus macaques were trained

173   to choose arrays that contained the greater ratio of positive to negative stimuli. Subjects´

174   performance was modulated by the ratio between ratios: they responded more quickly and

175   accurately the higher the ratio between ratios was, regardless of the absolute number of stimuli

176   within the arrays. Results of these two studies (Drucker et al., 2015; McCrink and Wynn, 2007)

177   suggest that the ability to discriminate ratios is a function of the ratio between the ratios to be

178   discriminated, similarly as the ability to discriminate absolute quantities is a function of the ratio

179   of the absolute set sizes. This raises the question whether the analogue magnitude system not only

180   enables an individual to *track* relative frequencies, but also to *use* relative frequency information

181   to draw statistical inferences from population to sample and vice versa.

182   One recent study did find some such evidence for an involvement of the analogue

183   magnitude system in decision making under uncertainty in nonhuman great apes: Hanus and Call

184   (2014) gave chimpanzees the choice between two trays on which food items were hidden under

185   cups. The trays differed with respect to the ratio of food items to cups and thus in chances of

186   finding food. Results showed that chimpanzees´ performance varied as a function of the ratio

187 between the two ratios, even in conditions where one tray constituted a 100% likelihood of finding

188 food. This suggests that the ratio between ratios, more than the magnitude of difference within the

189 single ratios, is decisive for apes´ ability to discriminate probabilities. However, chimpanzees´

190 success in this study could be explained with an absolute quantity heuristic, not regarding the

191 amount of food items available, but regarding the number of cups on each tray. In particular, the

192 tray depicting the more favorable food/cup ratio always held the smaller number of cups. The

193 authors added a control condition in which they excluded simple associative learning explanations

194 (subjects did not preferentially choose the tray with fewer cups when the food was visibly removed

195 from all cups). Nevertheless, it cannot be excluded that, as soon as there was any food to be found,

196 chimpanzees at least partially relied on a mental shortcut such as "fewer cups= higher likelihood

197 of finding food".

198   In sum, nonhuman great apes share with humans the fundamental ability to draw statistical

199 inferences from population to sample and (to a certain extent) vice versa. Yet, it still remains an

200 open question what the cognitive foundations of intuitive statistics are and whether they are the

201 same in humans and their closest living relatives. To date, two main reasons suggest that the

202 analogue magnitude system is the most plausible candidate for a basic statistical inference

203 mechanism. First, the prerequisite capacity for this kind of inferences, tracking ratios, shows the

204 same signatures as absolute quantity discrimination, both in human and nonhuman primates.

205 Second, decision making under uncertainty seems to be ratio dependent in chimpanzees. What is

206 missing is a comparative study testing great apes and humans in an intuitive statistical reasoning

207 task that systematically varies the ROR and, crucially, controls for the use of absolute quantity

208 heuristics.

209   The rationale of the present study, therefore, was threefold. First, we wanted to elucidate

210 the cognitive and evolutionary underpinnings of intuitive statistics in chimpanzees and humans by

211 testing its signatures. If intuitive statistics are based on the same analogue magnitude system, we

212 would expect the characteristic performance patterns and signatures in both species. Much like the

213 discrimination of absolute set sizes varies as a function of the ratio of the set sizes, we expected

214 that the accuracy of intuitive statistics would vary as a function of the ROR between sets. We were

215 particularly interested in determining the ROR lower threshold where performance breaks down

216 in each species. Moreover, a comparison of these signature limits with those found in simple

217 quantity discrimination tasks helped us determine whether the inclusion of the statistical operation

218 adds error to the representation in comparison to basic quantity discrimination. As described

219 earlier, quantities are represented in an analogue, approximate way. We assumed that

220 discriminating ratios of quantities *and* forming probabilistic expectations on their basis adds

221 considerable noise relative to absolute quantity discrimination, since it requires representing and

222 operating on quantities over multiple accounts (see, e.g. Barth et al. 2006 for an example of how

223 subtraction operations add error to the quantity representation in comparison to simple quantity

224 discrimination tasks). Accordingly, we expected the ROR threshold to be higher (i.e. less sensitive)

225 than the threshold for discriminating absolute quantities. To address these questions, we presented

226 chimpanzees with a previously established paradigm (Rakoczy et al., 2014) that required them to

227 reason from populations of food items with different ratios of preferred and non-preferred food

228 items to randomly drawn samples. We systematically varied the ratio between the two ratios (of

229 preferred to non-preferred food items; ROR) ranging from 1 (equal proportions in both

230 populations) to 16 (high magnitude of difference between populations). We tested human adults

231 in a computerized version of the same task.

232   Our second goal was to replicate and validate previous findings on intuitive statistics in

233   great apes. So far, only one experimentally highly experienced population of chimpanzees (and

234   other apes) has been tested for their statistical abilities (Rakoczy et al. 2014). Therefore, it remains

235   an open question whether findings of this particular population are generalizable to chimpanzees

236   as a species. To investigate whether intuitive statistics is in fact a common, natural capacity in

237   chimpanzees, we used the same task setup to test completely naïve individuals who were raised

238   and housed in a different environment. This allowed us to directly compare between the previously

239   tested, captive born and zoo housed chimpanzees, and the wild born, sanctuary housed

240   chimpanzees tested in the present study.

241   Lastly, our study rules out alternative explanations based on absolute quantity heuristics.

242   Most importantly, we included a crucial experiment to control for the possibility that apes used a

243   strategy based on avoiding the population with the largest number of non-preferred items. The to

244   date only study on nonhuman primates addressing such an avoidance heuristic has produced

245   negative results: Capuchin monkeys failed to choose the sample from the proportion wise

246   favorable population when they could not rely on the absolute amount of non-preferred items

247   (Tecwyn, Denison, Messer, & Buchsbaum, 2016). It is, therefore, crucial to explore the possibility

248   that apes´ success in the present and previous studies was due to this simple quantity heuristic.

249   Moreover, our study design ensured that apes could not succeed by choosing based on the absolute

250   quantity of preferred food items. While previous studies (Eckert et al. 2017; Rakoczy et al., 2014)

251   addressed this issue in separate control conditions, we designed our study in a way that apes were

252   prevented from using such a strategy in *all* test conditions. Lastly, we also controlled for heuristics

253   dealing with the total amount of food in each population. Hence, this is the first study on intuitive

254 statistics in great apes comprehensively controlling for absolute number heuristics both regarding

255 preferred and non-preferred items as well as their absolute total amount.

2. Methods

257 2.1. Chimpanzees

258 2.1.1. Subjects

259 We tested 24 chimpanzees (12 females) aged between 9 and 32 years at Ngamba Island

260 Chimpanzee Sanctuary, Uganda. Research strictly adhered to the legal requirements of Uganda

261 and was approved and reviewed by the Ugandan Wildlife Authorities and the Ugandan National

262 Council for Science and Technology. The study was approved by the ethics committees of the Max

263 Planck Institute for Evolutionary Anthropology and the Chimpanzee Sanctuary & Wildlife

264 Conservation Trust.

265 2.1.2. Design and procedure

266 Subjects were tested in eight ROR conditions and one condition controlling for the usage

267 of a non-preferred food avoidance heuristic (hereafter: carrot avoidance control; see Fig 2). All

268 conditions consisted of 12 trials, presented in a single session. The sequence of conditions was

269 randomized for each subject, with the exception of the carrot avoidance control, which was the

270 fifth condition for all subjects (see SI for more details).

271 We presented subjects with two transparent buckets filled with mixed populations of

272 peanuts and carrot pieces of roughly equal size and shape. In all but one condition (ROR 1), one

273 of the buckets contained a population that was more favorable in terms of its proportion of peanuts

274 to carrots compared to the other. The experimenter showed both buckets to the subject, directed

275 her gaze towards the ceiling and drew one item (always of the majority type) out of each of the

276 buckets in a way that the chimpanzee could not see what was drawn. The experimenter kept the

277 items hidden in her fists and the subject was allowed to indicate a choice to receive the chosen

278 sample. In half of the trials, the experimenter crossed her arms when moving the fists towards the

279 mesh to ensure that subjects made a choice between samples and not just chose the side where the

280 favorable population was still visible. Trials with and without crossing were alternated (see Fig 1

281 for an illustration).

282

283

284
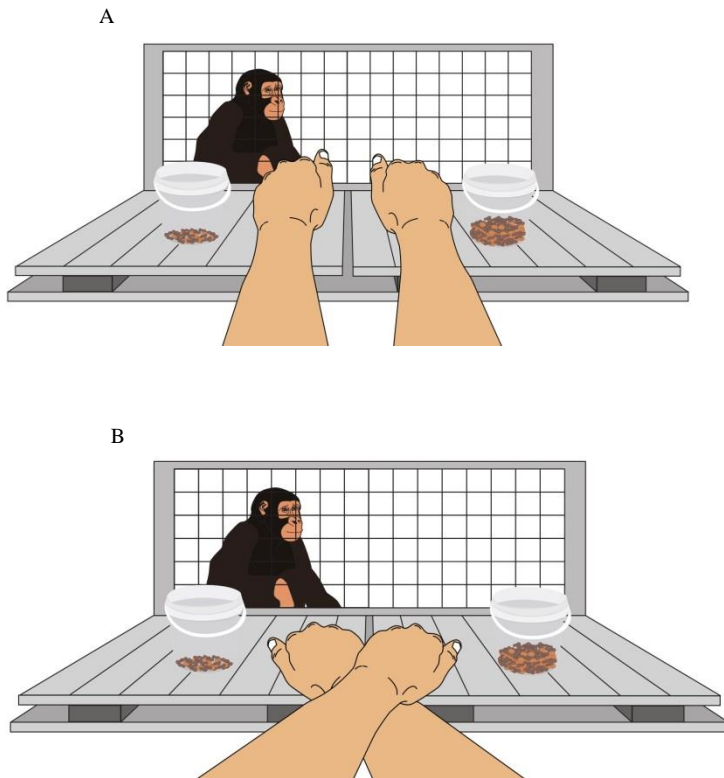
285 **Fig 1 Experimental setup. The experimenter simultaneously drew one sample from each of two populations in a**
286 **way that kept the object hidden from the chimpanzee. The experimenter then gave the subject a choice between**
287 **the two hidden samples (A). In half of the trials the experimenter crossed her arms before the subject was given a**
288 **choice (B).**

289

2.1.2.1. ROR conditions

To find the signatures of intuitive statistics, we systematically varied the ratio between populations` ratios (ROR) ranging from 1 (equal ratio in both populations) to 16 (high magnitude of difference between ratios). In all ROR conditions we disentangled absolute and relative frequencies of peanuts: the bucket containing the less favorable ratio of peanuts to carrots contained twice as many peanuts as the bucket containing the more favorable ratio (see Fig 2). This ensured that subjects truly reasoned about proportional information, rather than about absolute numbers of peanuts. RORs were calculated in the following way (following Drucker et al., 2016; for more details see SI):

$$\frac{\text{Ratio of peanuts to carrots in the more favorable population}}{\text{Ratio of peanuts to carrots in the less favorable population}}$$

One consequence of this was that the favorable population always contained a smaller total amount of food items across ROR conditions (see Fig 2). Hence, it is theoretically possible that subjects learned over the course of sessions to always pick the sample from the bucket with fewer items. Although this seems unlikely considering that chimpanzees have proven to be poor in learning arbitrary associations (e.g., Call, 2004), we included a ROR 1 condition in which both populations contained the same proportion of peanuts to carrots (i.e. both of them were equally likely to lead to a peanut as randomly drawn sample), while one contained double the amount of food items in absolute terms. If they still preferred the sample of one of the two populations, this would suggest that chimpanzees used other information than intended to solve the task. The reward pattern in this condition was adjusted accordingly: In randomized order, the experimenter drew a peanut from both buckets in half of the trials; in the other half she drew a carrot.

15

311   *2.1.2.2. Carrot avoidance control*

312   To test whether subjects solved the task using a non-preferred food avoidance strategy

313   focusing on the absolute number of carrots, here the favorable population contained four times

314   more carrot pieces than the unfavorable population, while the ROR was 16 (see Fig 2). If

315   chimpanzees used a carrot avoidance strategy, we expected them to perform worse than in the

316   ROR 16 condition.

ROR		1	1,5	2	4	6	8	12	16	Carrot avoidance
Proportions	Favorable population	28:28	28:23	28:20	28:14	28:11	28:10	28:8	28:7	320:80
	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.
	Unfavorable population	56:56	56:69	56:80	56:112	56:132	56:160	56:192	56:224	5:20
Likelihood of peanut sample	Favorable population	50%	54.9%	58.3%	66.7%	71.8%	73.7%	77.7%	80%	80%
	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.
	Unfavorable population	50%	44.8%	41.2%	33.3%	29.8%	25.9%	22.6%	20%	20%

317

318   **Fig 2 List of conditions (eight ROR conditions and carrot avoidance control) and the respective proportions within**
319   **the populations as well as the likelihood of drawing a peanut as a sample. Numerals in front of the colon depict**
320   **numbers of preferred items, numerals after the colon depict numbers of non-preferred items. The first line always**
321   **displays the favorable population except in ROR 1 in which both proportions were identical.**

322

323   2.1.3. Coding and data analysis

324   The apes´ choice was coded live by the experimenter. A second blind observer coded 25%

325   of the trials from video. Both raters were in excellent agreement ($K = .95$, $N = 576$). To investigate

326   whether there was an effect of the ROR on chimpanzees' performance across ROR conditions, we

327 ran a Generalized Linear Mixed Model (GLMM; Baayen 2008). "Correct choice" (choice of

328 sample from population with higher peanut proportion) was the dependent variable. Note that data

329 for the ROR 1 condition was not included in the model, since there was no "correct choice" in this

330 condition (both populations were equally likely to lead to a preferred food item as a reward). As

331 fixed effects we included logROR (since we expected the effect to be logarithmic), session and

332 trial number (to test for potential learning effects) as well as all second and third order interactions

333 among logROR, session number and trial number. To control for the effect of age and age^2 (in case

334 of a nonlinear age-effect) they were included as further fixed effects. Subject ID and session ID

335 were included as random effects. To keep type I error rate at the nominal level of 5% (Schielzeth

336 & Forstmeier 2009; Barr 2013) we included all possible random slopes components (logROR,

337 session number, trial number within subject ID and trial number within session ID) and also the

338 respective correlations between random slopes and intercepts (see SI for more details on the

339 statistical analysis). In order to determine the ROR lower threshold where performance breaks

340 down, we inspected the confidence interval limits of the model: The model predicts performance

341 to be above chance level in conditions for which the confidence interval limits lie above 0.5.

342   To investigate whether chimpanzees' performance in the carrot avoidance control was

343 different from the ROR 16 condition, we ran a second GLMM. Again, "correct choice" was the

344 response variable. As fixed effects we included condition (carrot avoidance control vs. ROR 16),

345 session number and trial number as well as the two-way interactions between condition and trial

346 number and between session number and trial number. Again, we included age and age^2 as further

347 fixed effects, subject ID and session ID as random effects, and all random slopes components

348 (condition, session number and trial number within subject ID and trial within session ID) as well

349 as the respective correlations between random slopes and intercepts (see SI for more details).

350   2.2. Humans

351   2.2.1. Subjects

352   We tested 144 adult humans (80 women) aged between 18 and 34 years at the University

353   of Göttingen, Germany. Participants were tested in a computerized version of the same task as the

354   chimpanzees. On a test computer, they were invited to imagine collecting as many red balls as

355   possible from pairs of transparent urns filled with red and blue balls. They were asked to envision

356   drawing from one urn of each pair with eyes closed and to indicate their choice by pressing one of

357   two keys as quickly as possible.

358   2.2.2. Design and procedure

359   Similar to the apes, humans were tested in eight ROR conditions ranging from 1 to 16, with

360   proportions resembling those for chimpanzees (see Fig 2). To prevent participants from learning

361   to always pick the urn with the smaller absolute quantity (since humans were tested in a single

362   session the likelihood for learning such a rule was high), we also tested seven additional conditions

363   (RORs between 1.5 and 16) in which absolute and relative frequencies were confounded. Since

364   we expected getting a ceiling effect for confounded conditions (and their interpretation would not

365   have been meaningful in terms of statistical reasoning), we did not plan to focus on these

366   conditions in the analysis (but see SI for results). To prevent participants from counting the balls,

367   images were only displayed for a maximum of 4 seconds (see SI for an example stimulus).

368   Participants saw six trials per condition and trials of all conditions were randomized for each

369   subject in one single test session.

370   2.2.3. Coding and data analysis

371   Participants´ choice and response time were recorded automatically by EPrime (mean

372   response times are depicted in SI Fig. 2). We used the same analysis as for chimpanzees, with the

373   following exceptions: We ran two separate models, one for disentangled RORs and one for

374   confounded RORs. Since humans were tested in a single session, we did not include session

375   number as fixed effect, nor did we include session ID as random effect. Further, we only included

376   age, but not age^2 as fixed effect, since we tested a small age range and did therefore not expect a

377   nonlinear effect. Due to a significant effect of trial number, we ran the model for disentangled

378   RORs again with only trial 1 performance considered (see SI for further details and results of the

379   trial 1 model).

380   **3. Results**

381   3.1. Chimpanzees

382   3.1.1. ROR conditions

383   Chimpanzees performance was significantly influenced by the logROR (GLMM,

384   estimate±SE=0.21±0.05, X^2=15.44, df=1, P<0.001; see Table 1 and SI for details), i.e. performance

385   increased as a function of the ROR (see Fig 3 A) from a mean of 56.9% correct trials in ROR 1.5

386   to 69.8% in ROR 16 (see Table 2). The mean number of correct trials in ROR 1 was 51.4%. The

387   model predicted the limit of chimpanzees´ abilities to be between ROR 2 and 4 (see Fig 3 A).

388   There was no effect of session or trial number, suggesting that chimpanzees' performance did not

389   change with increasing experience within a session or over the course of sessions (see Table 1 and

390   SI for more detailed information).

Table 1 Influence of logROR, session, trial, age and age² on chimpanzees´ proportion of correct choices

Term	Estimate	SE	X^2	Df	P
Intercept	0.60	0.11	(1)	(1)	(1)
logROR[2]	0.21	0.05	15.44	1	p<0.001
session[2]	0.12	0.07	2.65	1	0.098
trial[2]	-0.02	0.05	0.19	1	0.666
age[2]	0.03	0.09	0.10	1	0.749
(age)²[2]	-0.08	0.06	2.06	1	0.143

Note: [1] **not shown because lacking a meaningful interpretation;** [2] **these predictors were z-transformed**

3.1.2. Carrot avoidance control

We found that subjects performed significantly better in the carrot avoidance control compared to the ROR 16 condition (GLMM, estimate±SE=0.78±0.25, X^2=9.44, df=1, P=0.0016; see Fig 3 A and SI for details), which suggests that they did not use a strategy based on avoiding the population with more non-preferred food items.

Table 2 Mean percentage of correct choices for each condition. *In ROR1 was no correct answer; here the percentage depicts the mean proportion of trials in which subjects chose the sample drawn from the population with the smaller absolute number of items.

Species		ROR1	ROR1.5	ROR2	ROR4	ROR6	ROR8	ROR12	ROR16	Carrot Avoidance
Chimpanzees	Mean proportion correct choices	51.4%*	56.9%	59.0%	53.1%	63.4%	66.2%	64.2%	69.8%	81.9%
Humans	Mean proportion correct choices	51.9%*	43.4%	59.8%	69.3%	72.1%	78.6%	79.5%	80.0%	/

403
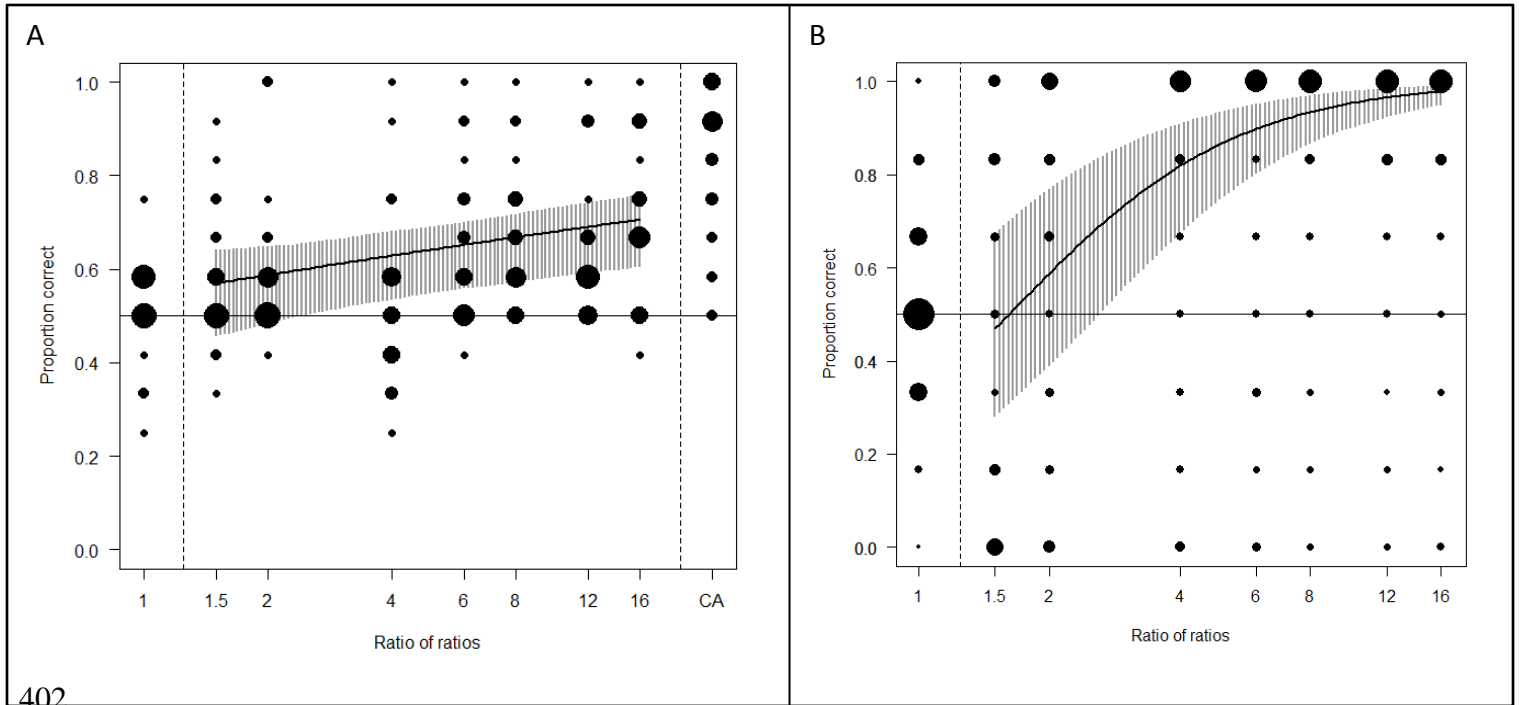
Fig 3 Chimpanzee (A) and human (B) performance across conditions. Shown is the proportion of trials in which subjects chose the sample from the favourable population for all (disentangled) ROR conditions (in ROR 1 proportion of trials in which subjects chose sample from smaller population) and, for chimpanzees the carrot avoidance control (CA). The dot size indicates the number of subjects performing at the same level (for readability purposes we used a different scale for the number of subjects in humans and chimpanzees). The horizontal line depicts chance level. The solid diagonal and vertical lines indicate the fitted model and its confidence limits. The model predicts performance to be above chance level in conditions for which the confidence interval limits lie above 0.5.

411

3.2. Humans

Humans´ performance was significantly influenced by the logROR (GLMM, estimate±SE=1.36±0.08, X^2=163.53, df=1, P<0.001; see Table 3 and SI for details): Performance increased logarithmically as a function of the ROR (see Fig 3 B) from a mean of 43.4% correct trials in ROR 1.5 to 80.0% in ROR 16 (see Table 2 and Fig 3B). The model predicted the same limit as for chimpanzees. We found an effect of trial number (X^2=4.58, df=1, P=0.029) indicating that participants´ performance slightly increased with increasing experience. However, significance of the overall results did not change when considering trial 1 performance only (see SI for more information).

421 **Table 3 Influence of logROR, trial number and age and on humans´ proportion of correct choices**

Term	Estimate	SE	X^2	df	P
intercept	1.96	0.27	(1)	(1)	(1)
logROR[2]	1.36	0.08	163.53	1	p<0.001
trial[2]	0.16	0.07	4.59	1	0.029
age[2]	-0.32	0.27	1.36	1	0.239

422 Note: [1] not shown because lacking a meaningful interpretation; [2] these predictors were z-transformed

4. Discussion

424   We found that chimpanzee and human performance in a task requiring inferences from

425 population to sample varied systematically as a function of the ROR, i.e., the magnitude of

426 difference between the ratios of two populations. In accordance with Weber´s law, performance

427 increased logarithmically as a function of the ROR. This is the first piece of evidence to suggest

428 that the analogue magnitude system is involved in intuitive statistical reasoning in both species.

429 Intriguingly, and although methods for both species were somewhat different (e.g. live

430 demonstration for chimpanzees vs. computer setup for humans) chimpanzees and humans

431 displayed the same approximate limit: While the model predicted above chance level performance

432 for ratios that differed by a factor of 4, it predicted failure for those that differed by a factor ≤2.

433 Given that this is the first study addressing this topic and considering the small sample size, these

434 results should be treated with caution. Nonetheless, the rather high threshold may suggest that, in

435 comparison to simple quantity discrimination tasks, the statistical operation adds some error to the

436 representation. Human adults, for instance, are able to discriminate absolute set sizes that differ by

437 a factor of 1.15 (e.g. Barth, Kanwisher, and Spelke, 2003; Pica, Lemer, Izard, & Dehaene, 2004)

438 compared to 6-month old infants who can discriminate ratios >1.5 (Lipton & Spelke, 2003; Xu &

439 Spelke, 2000). Not many studies documented the threshold for absolute quantity discrimination in

440 nonhuman primates. Reported limits range from values as low as 0.9 for great apes (Hanus and

441 Call, 2007) to 1.25 for rhesus macaques (Brannon & Terrace, 2000). The finding of a higher

442 threshold for ratio discrimination within the realms of statistical inferences relative to basic

443 quantity discrimination is consistent with the idea that additive error is to be expected when an

444 organism represents and operates over multiple amounts (see, e.g. Barth et al., 2006 and McCrink

445 & Wynn, 2007 for discussion of this hypothesis). Future studies with other populations of

446 chimpanzees and humans, as well as different absolute quantities will have to examine whether

447 we are truly dealing with a universal signature limit of statistical reasoning abilities. While

448 chimpanzee performance was still far from ceiling even at the highest tested RORs, human

449 performance rapidly increased and reached a plateau at ROR 8. This resembles findings on

450 absolute quantity discrimination, where humans showed higher accuracy compared to other

451 primates (e.g. Cantlon & Brannon, 2007), therefore suggesting that the same cognitive mechanism

452 is utilized in both types of tasks. Future studies should test chimpanzees with a wider range of

453 RORs to investigate whether (and when) they, like humans, also reach a maximum performance

454 plateau.

455   One question that the present findings raise is whether intuitive statistics is based on an

456 analogue magnitude system that is potentially not restricted to numerosity, but extends to all kinds

457 of magnitudes, or alternatively on an approximate number system (ANS) in the more narrow sense,

458 restricted to numerosity alone. This is related to the broader debate of whether numerical cognition

459 is a quintessential cognitive domain with a specialized cognitive mechanism or whether the ANS

460 is part of a broad domain in which all quantitative dimensions share computational mechanisms

461 (see, e.g. Cantlon, Platt, & Brannon, 2009 and Lourenco, 2015 for reviews). One way to inform

462 this debate is to develop an intuitive statistical inference task disentangling discrete from

463 continuous quantity information.

464   Another important question is whether the signatures of intuitive statistics and their limits

465   extend to differences across life-stages known to apply in the case of the ANS with regard to

466   absolute set sizes. More specifically, do we find better accuracy in older compared to younger

467   individuals (analogous to findings by Halberda & Feigenson, 2008)? And are early inter-individual

468   differences in accuracy predictive of later explicit statistical reasoning, as it is the case regarding

469   the ANS with respect to absolute set sizes and later mathematical achievement (Mazzocco,

470   Feigenson, & Halberda, 2011)?

471   Our study not only gives insights into the cognitive foundation of intuitive statistics, the

472   results also replicate those from our previous study (Rakoczy et al. 2014) in which we tested

473   chimpanzees (and other nonhuman great apes) with substantial experience with cognitive testing.

474   In that study, we had included a critical condition in which absolute and relative frequencies of

475   preferred items were disentangled while keeping the ROR at a value of 16 – similar as in the

476   present ROR 16 condition. While chimpanzees in Rakoczy et al. (2014) chose the sample from the

477   favourable population in 66% of trials, chimpanzees in our present study performed correct in

478   69.8% of trials, even slightly exceeding the previously tested individuals´ performance. Therefore,

479   our ROR 16 condition replicated previous findings with a new population of chimpanzees that was

480   completely naïve to this kind of task and was raised and housed in a different environment (wild

481   born and sanctuary housed vs. captive born and zoo housed). This suggests that intuitive statistical

482   reasoning is a natural capacity in chimpanzees and not restricted to a single population with

483   extensive experimental experience. Recent studies with two different monkey species (capuchin

484   monkeys: Tecwyn et al., 2016; long-tailed-macaques: Plací, Eckert, Rakoczy, and Fischer,

485   unpublished) using the same test paradigm failed to find unambiguous evidence for the presence

486   of intuitive statistical abilities in levels comparable to apes[3]. Although much more research is

487   needed, this might indicate that statistical reasoning has emerged late in primate evolution, perhaps

488   only in the ape lineage. It would be of great interest to investigate the ecological pressures that

489   could have led to the evolution from quantity discrimination abilities to probabilistic reasoning

490   capacities.

491   Importantly, and in contrast to previous studies, our test design comprehensively controlled

492   for the usage of simple heuristics based on absolute numbers of food items. In all ROR conditions

493   we disentangled absolute and relative frequencies of peanuts, excluding the possibility that

494   chimpanzees succeeded by simply picking the sample from the population with absolutely more

495   preferred items. Crucially, the carrot avoidance control, which had not been tested in previous

496   studies with apes, revealed that chimpanzees did not simply avoid the population containing more

497   non-preferred food items (by contrast, capuchin monkeys did not perform significantly above

498   chance level in a comparable condition in Tecwyn et al., 2016). Moreover, our ROR 1 condition

499   provided evidence that chimpanzees truly used proportional information to solve the task, rather

500   than, e.g. a "choose the sample from the bucket with less food"-strategy. Lastly, we did not detect

501   any effect of session or trial number on chimpanzees´ performance, making it unlikely that subjects

502   used strategies learned over trials.

503   In sum, this study revealed that the signatures of intuitive statistics in chimpanzees and

504   humans closely resemble those found in quantity discrimination tasks, thus strongly suggesting

---

[3] On the group level, long-tailed macaques were only successful in conditions in which they could rely on a quantity heuristic dealing with the absolute number of preferred food items (e.g. 64:16 vs. 16:64). They failed in conditions, in which absolute and relative frequencies were disentangled (e.g. 48:12 vs. 12:192; Placi et al., unpublished). Capuchin monkeys, by contrast, succeeded even in conditions disentangling absolute and relative frequencies of preferred food items. Their performance in a non-preferred food avoidance control, however, was not significantly different from chance, even when individuals who exhibited a side-bias in this condition were removed from the analysis (Tecwyn et al., 2016).

505 that these two abilities share the same basic and evolutionary ancient cognitive foundation, the

506 analogue magnitude system. Moreover, we replicated previous findings on statistical reasoning in

507 great apes with a new population of chimpanzees with a different housing and rearing background,

508 suggesting that intuitive statistics is in fact a common capacity in chimpanzees. Lastly, this is the

509 first study on intuitive statistics in great apes controlling for absolute number heuristics both

510 regarding preferred and non-preferred items as well as absolute total amount, providing further

511 evidence for true intuitive statistical reasoning in chimpanzees.

Acknowledgements

References

526 **References**

527   Ain, S. A., Giret, N., Grand, M., Kreutzer, M., & Bovet, D. (2009). The discrimination of

528 discrete and continuous amounts in African grey parrots (Psittacus erithacus). Anim Cogn, 12(1),

529 145-154. doi:10.1007/s10071-008-0178-8

530   Barnard, A. M., Hughes, K. D., Gerhardt, R. R., Divincenti, L., Jr., Bovee, J. M., &

531 Cantlon, J. F. (2013). Inherently Analog Quantity Representations in Olive Baboons (Papio

532 anubis). Front Psychol, 4, 253. doi:10.3389/fpsyg.2013.00253

533   Baayen, R. H. (2008). Analyzing Linguistic Data: A Practical Introduction to Statistics.

534 Cambridge: Cambridge University Press.

535   Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects

536 models. Front Psychol, 4, 328. doi:10.3389/fpsyg.2013.00328

537   Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number

538 representations in adults. Cognition, 86(3), 201-221.

539   Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-

540 symbolic arithmetic in adults and young children. Cognition, 98, 199–222.

541   Beran, M. J. (2004). Chimpanzees (Pan troglodytes) respond to nonvisible sets after one-

542 by-one addition and removal of items. J Comp Psychol, 118(1), 25-36. doi:10.1037/0735-

543 7036.118.1.25

544   Beran, M. J. (2017). Quantitative Cognition. In J. Call (Ed.), APA Handbook of

545   Comparative Psychology. Vol 2: Perception, learning, and cognition (pp. 535-577). Washington,

546   DC: American Psychological Association

547   Bortz, J. (1999). Statistik für Sozialwissenschaftler. Springer. Berlin, Heidelberg.

548   Brannon, E. M., & Terrace, H. S. (2000). Representation of the numerosities 1–9 by rhesus

549   macaques (Macaca mulatta). J of Exp Psychol: Anim Behav Process, 26(1), 31.

550   Buckingham, J. N., Wong, B. B. M., & Rosenthal, G. G. (2007). Shoaling decisions in

551   female swordtails: how do fish gauge group size? Behaviour, 144, 1333-1346.

552   doi:10.1163/156853907782418196

553   Call, J. (2000). Estimating and operating on discrete quantities in orangutans (Pongo

554   pygmaeus). Journal of Comparative Psychology, 114(2), 136-147. doi:Doi 10.1037//0735-

555   7036.114.2.136

556   Call, J. (2004). Inferences about the location of food in the great apes (Pan paniscus, Pan

557   troglodytes, Gorilla gorilla, and Pongo pygmaeus). J Comp Psychol, 118(2), 232-241.

558   doi:10.1037/0735-7036.118.2.232

559   Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large

560   numbers in monkeys and humans. Psychol Sci, 17(5), 401-406. doi:10.1111/j.1467-

561   9280.2006.01719.x

562   Cantlon, J. F., & Brannon, E. M. (2007). Basic math in monkeys and college students.

563   PLoS Biol, 5(12), e328. doi:10.1371/journal.pbio.0050328

564   Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. Trends

565   Cogn. Sci. (Regul. Ed.) 13, 83–9110.1016/j.tics.2008.11.007

566   Carey, S. (2009). The origins of concepts. New York: Oxford University Press.

567   Davidson, D. (1995). The representativeness heuristic and the conjunction fallacy effect in

568   children's decision making. Merrill-Palmer Quarterly (1982-), 328-346.

569   Dehaene, S. (2011). The number sense: How the mind creates mathematics. New York,

570   NY: Oxford University Press

571   De Neys, W., & Vanderputte, K. (2011). When less is not always more: Stereotype

572   knowledge and reasoning development. Dev Psychol, 47(2), 432.

573   Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very

574   young infants: evidence from 4.5- and 6-month-olds. Dev Psychol, 49(2), 243-249.

575   doi:10.1037/a0028278

576   Denison, S., Trikutam, P., & Xu, F. (2014). Probability versus representativeness in

577   infancy: can infants use naive physics to adjust population base rates in probabilistic inference?

578   Dev Psychol, 50(8), 2009-2019. doi:10.1037/a0037158

579   Denison, S., & Xu, F. (2010a). Twelve- to 14-month-old infants can predict single-event

580   probability with large set sizes. Dev Sci, 13(5), 798-803. doi:10.1111/j.1467-7687.2009.00943.x

581   Denison, S., & Xu, F. (2010b). Integrating physical constraints in statistical inference by

582   11-month-old infants. Cogn Sci, 34(5), 885-908. doi:10.1111/j.1551-6709.2010.01111.x

583   Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants.

584   Cognition, 130(3), 335-347.

585   Drucker, C. B., Rossa, M. A., & Brannon, E. M. (2016). Comparison of discrete ratios by

586   rhesus macaques (Macaca mulatta). Anim Cogn, 19(1), 75-89. doi:10.1007/s10071-015-0914-9

587   Eckert, J., Rakoczy, H., & Call, J. (2017). Are great apes able to reason from multi-item

588   samples to populations of food items? Am J Primatol. doi:10.1002/ajp.22693

589   Eckert, J., Rakoczy, H., Call, J., Herrmann, E., & Hanus, D. (2018). Chimpanzees Consider

590   Humans' Psychological States when Drawing Statistical Inferences. Curr Biol.

591   doi:https://doi.org/10.1016/j.cub.2018.04.077Halberda, J., & Feigenson, L. (2008).

592   Developmental change in the acuity of the 'number sense': the approximate number system in 3-,

593   4-, 5-, and 6-year-olds and adults. Developmental Psychology, 44 (5), 1457–1465.

594   Hanus, D., & Call, J. (2007). Discrete quantity judgments in the great apes (Pan paniscus,

595   Pan troglodytes, Gorilla gorilla, Pongo pygmaeus): the effect of presenting whole sets versus item-

596   by-item. J Comp Psychol, 121(3), 241-249. doi:10.1037/0735-7036.121.3.241

597   Hanus, D., & Call, J. (2014). When maths trumps logic: probabilistic judgements in

598   chimpanzees. Biol Lett, 10(12), 20140892. doi:10.1098/rsbl.2014.0892

599   Jacobs, J. E., & Potenza, M. (1991). The use of judgement heuristics to make social and

600   object decisions: A developmental perspective. Child Dev, 62(1), 166-178.

601   Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of

602   representativeness. Cog psych, 3(3), 430-454.

603   Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychol rev, 80(4),

604   237.

605   Kayhan, E., Gredebäck, G., & Lindskog, M. (2017). Infants distiguish between two events

606   based on their relative likelihood. Child Dev. doi: 10.1111/cdev.12970

607   Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number

608   discrimination in human infants. Psychol sci, 14(5), 396-401.

609   Lourenco, S. F. (2015). On the relation between numerical and non-numerical magnitudes:

610   Evidence for a General Magnitude System. In D.C. Geary, D.B. Berch, & K.M. Koepke (Eds.),

611   Mathematical cognition and learning. Vol 1: Evolutionary origins and early development of

612   number processing (pp. 145-174). London, San Diego, Waltham and Oxford: Elsevier

613   Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the

614   approximate number system underlies mathematical learning disability. Child Dev. 82, 1224–1237

615   10.1111/j.1467-8624.2011.01608.x

616   McCrink, K., & Wynn, K. (2007). Ratio abstraction by 6-month-old infants. Psychol Sci,

617   18(8), 740-745. doi:10.1111/j.1467-9280.2007.01969.x

618   Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. Annual Review

619   of Neuroscience, 32, 185-208. doi:10.1146/annurev.neuro.051508.135550

620   O´Grady, S., Griffiths, T. L., Xu, F. (2016). Do simple probability judgements rely on

621   integer approximation? Proceedings of the 38th Annual Conference of the Cognitive Science

622   Society

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

624 Leake et al.).

625   Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in

626 an Amazonian Indigene Group. Science, 306, 499–503.

627   Rakoczy, H., Cluver, A., Saucke, L., Stoffregen, N., Grabener, A., Migura, J., & Call, J.

628 (2014). Apes are intuitive statisticians. Cognition, 131(1), 60-68.

629 doi:10.1016/j.cognition.2013.12.011

630   Rugani, R., Cavazzana, A., Vallortigara, G., & Regolin, L. (2013). One, two, three, four,

631 or is there something more? Numerical discrimination in day-old domestic chicks. Anim Cogn,

632 16(4), 557-564. doi:10.1007/s10071-012-0593-8

633   Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident

634 estimates in mixed models. Behav Ecol, 20(2), 416-420. doi:10.1093/beheco/arn145

635   Tecwyn, E. C., Denison, S., Messer, E. J., & Buchsbaum, D. (2016). Intuitive probabilistic

636 inference in capuchin monkeys. Anim Cogn. doi:10.1007/s10071-016-1043-9

637   Téglás, E., Ibanez-Lillo, A., Costa, A., & Bonatti, L. L. (2015). Numerical representations

638 and intuitions of probabilities at 12 months. Developmental Science, 18, 183–193.

639 https://doi.org/10.1111/desc.12196

640   Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011).

641 Pure reasoning in 12-month-old infants as probabilistic inference. Science, 332(6033), 1054-1059.

642   Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases.

643   Science, 185(4157), 1124–1131.

644   Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of

645   choice. Science, 211(4481), 453–458.

646   Underwood, A. J. (1997). Experiments in Ecology. Cambridge University Press.

647   Cambridge.

648   Xu, F., & Denison, S. (2009). Statistical inference and sensitivity to sampling in 11-month-

649   old infants. Cognition, 112(1), 97-104. doi:10.1016/j.cognition.2009.04.006

650   Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. Proc Natl Acad Sci

651   U S A, 105(13), 5012-5015. doi:10.1073/pnas.0704450105

652   Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants.

653   Cognition, 74(1), B1-B11.

654   Zar, J. H. (1999). Biostatistical Analysis (4th ed.). Prentice Hall. New Jersey.