



Heriot-Watt University
Research Gateway

Intuitive probabilistic inference in capuchin monkeys

Citation for published version:

Tecwyn, EC, Denison, S, Messer, EJE & Buchsbaum, D 2016, 'Intuitive probabilistic inference in capuchin monkeys', *Animal Cognition*. <https://doi.org/10.1007/s10071-016-1043-9>

Digital Object Identifier (DOI):

[10.1007/s10071-016-1043-9](https://doi.org/10.1007/s10071-016-1043-9)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

Animal Cognition

Publisher Rights Statement:

The final publication is available at Springer via <http://dx.doi.org/10.1007/s10071-016-1043-9>

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Intuitive probabilistic inference in capuchin monkeys

Emma C Tecwyn¹, Stephanie Denison², Emily JE Messer¹, Daphna Buchsbaum^{3,1}

¹School of Psychology and Neuroscience, University of St Andrews, Scotland, UK

²Department of Psychology, University of Waterloo, Canada

³Department of Psychology, University of Toronto, Canada

Author Note

This research was supported by an Economic and Social Research Council Future Research Leaders grant to Daphna Buchsbaum.

Correspondence concerning this article should be addressed to Emma Tecwyn, Department of Psychology, University of Toronto, 100 St George Street, Toronto, Ontario, M5S 3G3, Canada. Email: emma.tecwyn@utoronto.ca

Emma C. Tecwyn is now at Department of Psychology, University of Toronto, Canada
Emily J.E. Messer is now at Department of Psychology, School of Life Sciences, Heriot-Watt University, Scotland, UK

PROBABILISTIC INFERENCE IN CAPUCHINS

20 **Abstract**

21 The ability to reason about probabilities has ecological relevance for many species. Recent research has
22 shown that both preverbal infants and non-human great apes can make predictions about single-item
23 samples randomly drawn from populations by reasoning about proportions. To further explore the
24 evolutionary origins of this ability, we conducted the first investigation of probabilistic inference in a
25 monkey species (capuchins; *Sapajus* spp.). Across four experiments, capuchins ($N = 19$) were presented
26 with two populations of food items that differed in their relative distribution of preferred and non-
27 preferred items, such that one population was more likely to yield a preferred item. In each trial,
28 capuchins had to select between hidden single-item samples randomly drawn from each population. In
29 Experiment 1 each population was homogeneous so reasoning about proportions was not required;
30 Experiments 2-3 replicated previous probabilistic reasoning research with infants and apes; and
31 Experiment 4 was a novel condition untested in other species, providing an important extension to
32 previous work. Results revealed that at least some capuchins were able to make probabilistic inferences
33 via reasoning about proportions as opposed to simpler quantity heuristics. Performance was relatively
34 poor in Experiment 4, so the possibility remains that capuchins may use quantity-based heuristics in some
35 situations, though further work is required to confirm this. Interestingly, performance was not at ceiling in
36 Experiment 1, which did not involve reasoning about proportions, but did involve sampling. This suggests
37 that the sampling task posed demands in addition to reasoning about proportions, possibly related to
38 inhibitory control, working memory, and/or knowledge of object permanence.

39

40 *Keywords:* capuchin; intuitive statistics; numerical cognition; primate cognition; probabilistic inference;
41 proportional reasoning

PROBABILISTIC INFERENCE IN CAPUCHINS

42 **Introduction**

43 Numerical competence is ecologically relevant in many contexts. It enables efficient foraging,
44 reduces predation risk, increases the likelihood of success in group conflict situations, and makes it
45 possible to keep track of group members and prey items (e.g. Addessi et al. 2008; Beran et al. 2011;
46 Schmitt and Fischer, 2011; Wilson et al. 2001). Extensive research has revealed that basic numerical
47 abilities are evolutionarily ancient: a wide range of nonhuman animals (hereafter animals) including
48 several species of mammals, birds, fish and insects are capable of using representations of quantity to
49 guide their behaviour (see Reznikova and Ryabko 2011; and Vallortigara 2014 for recent reviews).

50 One specific aspect of numerical cognition that has been much less studied in animals is the
51 ability to reason about probabilities, or make probabilistic inferences. The key distinction between this
52 ability and other types of numerical competence is that reasoning about probabilities involves reasoning
53 about *relative* quantities, or proportions (e.g. in a population consisting of two types of item, the quantity
54 of one type of item *relative* to the total quantity of both types of item) as opposed to simple comparisons
55 of *absolute* quantities (Bryant and Nunes 2012). In some situations in the natural environment the ability
56 to make accurate absolute quantity judgements is not sufficient for informing decision-making; being able
57 to use proportion judgements is also required (Rugani et al. 2015). For example, to gain access to the
58 largest quantity of food, an individual needs to consider both the amount of food in alternative locations,
59 and the number of other individuals feeding at these different locations (Rugani et al. 2015). Relative
60 judgments are also important outside of the number domain: there is a growing literature on inequity
61 aversion in animals – the sensitivity to one’s own effort and payoff *relative* to another individual’s (e.g.
62 Brosnan and de Waal 2003; Brosnan et al. 2005; Cronin and Snowden 2008; Range et al. 2009).

63 In humans, traditional theory suggests that the ability to make probabilistic inferences does not
64 develop until around seven years of age (Piaget and Inhelder 1975). However, recent research using
65 violation of expectation looking-time paradigms (based on the premise that infants look longer at

PROBABILISTIC INFERENCE IN CAPUCHINS

66 surprising or unexpected events) and action-based choice tasks has revealed that human infants are
67 capable of basic reasoning about probabilities (Denison and Xu 2010; 2014; Teglas et al. 2007; 2011; Xu
68 and Garcia 2008). Denison and Xu (2010) also demonstrated that infants are capable of drawing
69 inferences from populations to randomly-drawn single-item samples to guide their decision-making in a
70 choice task. When presented with two visible populations that differed in their distributions of preferred
71 to non-preferred items (4:1 vs. 1:4), infants accurately predicted which of two single-item samples drawn
72 from the two populations was more likely to consist of a preferred item, indicated by crawling towards the
73 location of that sample.

74 In all of these studies however, absolute quantity was confounded with proportion, because in
75 each case the highest proportion object in a population was also the most numerous. For example,
76 suppose a person has a stronger preference for pink than green objects. When comparing a population
77 containing 40 pink and 10 green objects against a population of 10 pink and 40 green objects, one could
78 compare the two proportions (4:1 vs. 1:4) or one could use a shortcut and simply compare the quantity of
79 pink objects only across populations (40 vs. 10). Though correct use of either strategy is likely to result in
80 the same behaviour in this case (i.e. approach the sample from the 40 pink and 10 green population), only
81 the latter strategy reflects accurate probabilistic reasoning, as using a strategy of relying on numerators
82 and ignoring denominators will lead to errors in many cases. Indeed, ignoring denominators is a strategy
83 that children have been shown to use in some mathematics problems until middle childhood, and evidence
84 of proportional reasoning is required for a population to be credited with true probabilistic reasoning
85 (Falk et al. 2012; Bryant & Nunes, 2012).

86 To address this issue, Denison and Xu (2014) ran a series of experiments to determine whether
87 infants are using this type of quantity-based heuristic or comparing proportions when making inferences.
88 Infants were presented with two visible populations that differed in their distributions of preferred to non-
89 preferred items, as in Denison and Xu (2010). However in this series of experiments infants could not
90 succeed by basing their selection on the greater quantity of preferred items, because the quantity was the

PROBABILISTIC INFERENCE IN CAPUCHINS

91 same in both populations, or because quantity was pitted directly against proportion (i.e. the population
92 containing the greater quantity of preferred items contained a lower proportion of preferred items). Their
93 results provided strong evidence that 12-month-old infants are capable of using proportions to predict
94 which of two single-item samples randomly drawn from two populations is more likely to consist of a
95 preferred (as opposed to non-preferred) item.

96 In addition to investigating the developmental origins of probabilistic reasoning in Western
97 children, recent research has begun to explore this capacity cross-culturally, and has revealed that
98 preliterate and prenumerate human cultures are able to make implicit probabilistic inferences, suggesting
99 that this ability may be universal within our own species (Fontanari et al. 2014). There is also a growing
100 body of literature investigating the evolutionary origins of intuitive statistics; that is, the extent to which
101 any animals might share intuitive statistical abilities with humans. Rakoczy et al. (2014) ran a study based
102 on the tasks developed by Denison and Xu (2010; 2014) with all four species of nonhuman great ape
103 (hereafter ape). They found that apes share with human infants the ability to draw inferences from
104 populations to randomly drawn single-item samples. Several control conditions ruled out the possibility
105 that apes were solving the tasks by using simple quantity heuristics or subtle experimenter-given cues, as
106 opposed to reasoning about proportions (Rakoczy et al. 2014). Further evidence that apes are capable of
107 making basic probabilistic inferences comes from a study by Hanus and Call (2014), which investigated
108 chimpanzees' ability to use probabilistic reasoning to find a food item hidden under one of several cups
109 on one of two trays. Performance in the task was correlated with the probability ratio between the two
110 trays (a signature property of the analogue magnitude system (AMS); a mechanism for quantification of
111 arbitrarily large magnitudes that is shared by many species, e.g. Jordan and Brannon 2006), so the greater
112 the discrepancy between the two trays in terms of probability of finding the reward, the more likely
113 chimpanzees were to select a cup from the more probable tray.

114 These recent findings suggest that the capacity for probabilistic inference is not uniquely human;
115 rather it is shared by our closest relatives, the great apes. However, the question remains of how

PROBABILISTIC INFERENCE IN CAPUCHINS

116 evolutionarily ancient and therefore how widespread in the animal kingdom the ability may be. Recent
117 research has demonstrated that two individuals of an Old World monkey species (rhesus macaques;
118 *Macaca mulata*; Drucker et al. 2016), as well as day-old chicks (*Gallus gallus*; Rugani et al. 2016)
119 possess the pre-requisite ability of distinguishing between proportions of discrete items; however, these
120 studies did not address whether these individuals were able to make inferences on the basis of
121 probabilities. Probabilistic inference goes one step beyond the ability to compare proportions, because the
122 subject also needs to understand the sampling part of the procedure; that is, they need to make inferences
123 about the probable identity of items drawn from populations, based on the distribution of items in those
124 populations. The aim of the present set of experiments was to investigate whether capuchin monkeys
125 (*Sapajus* spp.), like human infants and apes, are able to use proportional reasoning to make probabilistic
126 inferences about single-item samples randomly drawn from populations. To our knowledge this is the first
127 study to investigate probabilistic inference in a monkey species. Capuchins are interesting from a
128 comparative perspective, because as a New World primate they share a more evolutionarily ancient
129 common ancestor with humans than the apes (and the Old World monkeys), the two lineages having
130 diverged over 30 million years ago (Fragaszy et al 2004). Previous research on numerical cognition in
131 capuchins has generally demonstrated that they have abilities comparable to those exhibited by apes. Like
132 apes, capuchins have displayed an ordinal concept of quantity (Judge et al. 2005); they are able to judge
133 relative quantity of sets of objects and amounts of substance when they are presented as discrete sets (e.g.
134 Addessi et al. 2008) and to some extent when they are presented sequentially (e.g. dropped into a cup one
135 item at a time; Evans et al. 2009; VanMarle et al. 2006); and they can make accurate numerosity
136 judgements when presented with moving dots of two different colours on a screen (Beran et al. 2011).
137 There is also some evidence that capuchins may be sensitive to inequity between themselves and another
138 individual (Brosnan and de Waal 2003), which also involves making relative judgements. We therefore
139 predicted that capuchins should also perform comparably with apes in proportional reasoning tasks.

PROBABILISTIC INFERENCE IN CAPUCHINS

140 We presented capuchins with a series of experiments based on those used in recent studies with
141 infants (Denison and Xu 2010; 2014), young children (Giroto et al. 2016) and apes (Rakoczy et al. 2014),
142 as well as a novel experimental condition that has not previously been presented to any species
143 (Experiment 4). In all of the experiments, subjects were presented with two populations of food items in
144 transparent jars. The two populations differed in terms of their distribution of two types of food item: one
145 preferred and one non-preferred; so that in each case one population was ‘favourable’, in terms of the
146 probability of a randomly drawn single-item sample consisting of a preferred item. Across all experiments
147 we refer to the favourable population as Jar A, and the unfavourable population as Jar B (though jar
148 placement is always counterbalanced on the left and right). In each trial, the experimenter randomly drew
149 a single-item sample from each jar, kept them hidden in her hands, and allowed the subject to choose
150 between the two samples. To select the sample most likely to consist of the preferred item, subjects had to
151 distinguish between the two populations and infer the relative probability that each of the samples would
152 consist of a preferred item. They then had to use this information to guide their decision-making
153 behaviour when selecting one of the samples. We also extended the recent work with infants and apes by
154 including a novel experimental condition (Experiment 4) that directly addressed the possibility that
155 subjects could potentially succeed at the task by using a quantity heuristic based on avoiding the
156 population containing the greater quantity of non-preferred items, by presenting populations that were
157 both unlikely to yield a preferred item, but one was more unlikely than the other.

158 Methods**159 Subjects**

160 Nineteen capuchins (*Sapajus* spp.) participated in this study (see Table 1). The subjects were
161 housed at the ‘Living Links to Human Evolution’ Research Centre at the Royal Zoological Society of
162 Scotland, Edinburgh Zoo, U.K. There were 6 females and 13 males aged between 2 and 16 years (mean =
163 6.4 years). The subjects did not have any previous experience with numerical or quantity based cognitive

PROBABILISTIC INFERENCE IN CAPUCHINS

164 tests. The subjects were housed in two groups (East and West; referring to the geographical location of
165 the enclosures at the zoo), and both groups cohabited with common squirrel monkeys (*Saimiri sciureus*).
166 Each group was housed in an indoor enclosure (189 m³) with access to a ~900 m² outdoor enclosure, both
167 of which had ample climbing substrates. For further details of housing and husbandry see Leonardi et al.
168 (2010).

169 ** Table 1 about here **

170 Study design

171 Prior to participating in any experiments, all subjects participated in food preference trials, to
172 establish their preference between a peanut and a monkey pellet. Subsequently, four experiments were
173 carried out (Experiments 1 – 4). Experiment 1 was designed to familiarise subjects with the single-item
174 sampling procedure, and establish their baseline performance in this task with two populations each
175 consisting of just one type of item (100% preferred vs. 100% non-preferred; Figure 1a). Therefore, all
176 subjects participated in Experiment 1 first.

177 ** Figure 1 about here **

178
179 Experiment 1 consisted of four sessions of six trials (24 trials in total): in sessions 1 – 3 the
180 experimenter's arms were straight (i.e. the hand containing the item from Jar A was next to Jar A when
181 the subject made their selection; Figure 2a), and in session 4 the experimenter crossed her arms before
182 allowing the subject to make their selection (i.e. the hand containing the item from Jar A was next to Jar B
183 when the subject made their selection; Figure 2b).

184 ** Figure 2 about here **

185
186 This design was used in the previous work with infants and apes thus we did the same to allow maximal
187 comparability across taxa. Including the arms-crossed session also allowed us to rule out the possibility
188 that capuchins were simply basing their selection on the location of the favourable population (e.g.

PROBABILISTIC INFERENCE IN CAPUCHINS

189 choosing the hand next to the jar where they could see the most preferred items, in which case we would
190 expect performance to be below chance in the arms-crossed session), or actually considering the samples
191 drawn from the populations. This is important because choosing on the basis of the probable identity of
192 the sample is an important way in which probabilistic inference differs from the pre-requisite ability of
193 being able to compare the proportions of items in populations. Without these arms-crossed trials, it would
194 be difficult to know whether the participants are truly reaching toward the correct sample or are instead
195 perhaps reaching toward the jar with the higher proportion of preferred items.

196 Experiments 2 – 4 were designed to investigate the ability of the subjects to make inferences
197 about random samples drawn from mixed populations (Figure 1b – d), and to rule out the possibility that
198 subjects could solve this type of problem using heuristic rules based on the *absolute quantities* of the
199 items, rather than the *relative proportions* of the preferred to non-preferred items. To control for potential
200 learning effects across experiments, subjects completed Experiments 2 – 4 in a random order.

201 Experiments 2 – 4 each consisted of three sessions of six trials (18 trials in total). Within each session, all
202 trials were either presented with the experimenter’s arms straight (Figure 1a) or crossed (Figure 1b).

203 Within each experiment subjects were randomly assigned to either arms straight or arms crossed
204 presentation, and across Experiments 2 – 4 subjects either experienced two experiments with arms
205 straight and one experiment with arms crossed, or vice versa (see Table S1 in Online Resource 1). To
206 control for side preferences, in all experiments, the side on which the jar containing the favourable
207 population (Jar A) was presented was pseudorandomised within each session of six trials, with the
208 constraints that it appeared three times on each side, and not on the same side in more than two
209 consecutive trials.

210 **Procedure and materials**

211 Subjects were tested individually in a test cubicle (49.5 cm × 52.1 cm × 51.4 cm) with a Plexiglas
212 window that had two 5 cm diameter holes 26 cm apart that subjects could reach their arms out of to make

PROBABILISTIC INFERENCE IN CAPUCHINS

213 selections. Subjects received one session of six trials per session and up to two sessions per day (with
214 approximately three hours between the morning and afternoon sessions). Populations of peanuts and
215 monkey pellets (Figure 1) were presented to subjects in two transparent glass jars on a wheeling trolley.

216 In all experiments, several measures were taken to avoid possible cueing via the experimenter's
217 body posture, facial expression or gaze direction (i.e. a "Clever Hans effect"). The general method for
218 drawing samples from populations and presenting them to subjects followed Rakoczy et al. (2014). At the
219 start of each trial the experimenter placed her closed fists on the table behind the two jars. She then
220 simultaneously shook both jars whilst looking at the subject to draw its attention to them. The
221 experimenter then closed her eyes and tilted her head upwards to convey random drawing of samples,
222 drew a single item from each jar simultaneously, and kept them concealed from the subject in her closed
223 fists (in fact the required items were already surreptitiously held in the experimenter's hands prior to
224 shaking the jars and "extracting" the sample). Following Rakoczy et al. (2014), in Experiments 2 and 3
225 the item "drawn" from each population was the majority item. Because in Experiment 4 the non-preferred
226 item was in the majority in both populations, we manipulated the samples to match the probabilities of the
227 populations (see procedure section of Experiment 4 for details). The experimenter then extended her
228 arms simultaneously to present her fists containing the concealed items centred at the two evenly spaced
229 holes in the cubicle window, at a fixed equal distance from the window. She then held this position until
230 the subject made their choice. When presenting items with arms crossed (Figure 2b), the experimenter
231 always crossed her right arm over her left arm. During presentation of the items the experimenter fixed
232 her gaze in the centre of the two holes in the window and maintained a symmetrical posture and neutral
233 expression (as in e.g. Albiach-Serrano and Call 2014). This prevented the experimenter from
234 inadvertently gazing at either option or making eye contact with the subject (it was not possible for the
235 experimenter to have her eyes closed or avert her gaze while the subject made their selection for safety
236 reasons). The subject was allowed to select one fist by touching it and the experimenter then opened that
237 hand and allowed the subject to take that item for immediate consumption. Subjects were not praised for

PROBABILISTIC INFERENCE IN CAPUCHINS

238 selecting either item. The experimental procedure can be seen in the supplementary videos (Online
239 Resources 2, 5, 7 and 8). Any deviations from this general procedure are described under the relevant
240 experiment section below.

241 **Data coding and analysis**

242 All sessions were videotaped. For each trial we scored whether the subject selected the hand
243 containing the item from Jar A (favourable population) or the hand containing the item from Jar B
244 unfavourable population). To select a hand the subject had to touch it with one of their hands; just
245 reaching towards one of the experimenter's hands did not constitute making a selection. We also scored
246 the side at which the hand that was selected by the subject was located (left or right window hole, from
247 the subject's perspective). A second coder scored a random 25% of the recorded sessions to assess inter-
248 observer reliability. Cohen's kappa was 0.99 for whether the subject selected the experimenter's hand
249 containing the item from Jar A or Jar B (99% agreement between coders). Disagreements were resolved
250 through discussion. Our main dependent variable was the average proportion of trials correct. We also
251 examined Trial 1 performance for each experiment, as well as Trial 1 performance for each session of
252 each experiment. The reason for this latter analysis was to increase power, given that we had fewer
253 subjects than the previous work with infants and apes. All statistical tests were two-tailed, and the
254 significance level of alpha was 0.05 unless otherwise stated.

255 **Preference trials**

256 Prior to introducing the populations of items in jars, food preference testing was carried out. The
257 aim of this was to establish each subject's preference between a peanut and a similar-sized monkey pellet
258 piece.

259 **Subjects**

260 All 19 subjects participated in the preference trials.

PROBABILISTIC INFERENCE IN CAPUCHINS

261 **Procedure**

262 Subjects were presented with a single session of 10 preference trials. In each trial, the
263 experimenter presented the two items simultaneously in her open palms at the left and right holes in the
264 cubicle window and the subject was allowed to take one item. The side on which the peanut appeared (left
265 vs. right window hole) was pseudorandomised, with the constraints that it appeared five times on each
266 side, and it could not appear on the same side in more than two consecutive trials.

267 **Results and discussion**

268 In the preference test all 19 subjects selected the peanut in 10/10 trials. This suggests that all of
269 the subjects had a strong preference for peanuts over monkey pellets, and were thus highly motivated to
270 maximise intake of peanuts. It also demonstrated that subjects were able to visually discriminate between
271 the two food items.

272 **Experiment 1: Inferences from homogeneous populations to samples (baseline condition)**

273 The aim of Experiment 1 was to familiarise subjects to the sampling procedure, and to establish
274 their baseline performance in the task when each of the populations consisted of a single type of item
275 (100% preferred vs. 100% non-preferred), i.e. when no proportional reasoning was necessary.

276 **Subjects**

277 All 19 subjects participated in Experiment 1 (see Table 1).

278 **Apparatus and procedure**

279 The jars depicted in Figure 1a were used. Jar A contained 300 peanuts (preferred) and Jar B
280 contained 300 pellets (non-preferred); i.e. the populations were not mixed and each consisted of one type
281 of item.

PROBABILISTIC INFERENCE IN CAPUCHINS

282 There were four sessions of six trials (24 trials in total). In session 1 (arms straight; Figure 2a)
 283 items were presented to the subject in the experimenter’s closed fists, and once the subject had selected a
 284 hand the experimenter opened that hand and the subject could take the item of food from their palm. The
 285 item concealed in the unselected hand was not revealed to the subject. Because performance was not as
 286 good as we might have expected in this initial session we made some modifications to this procedure. In
 287 sessions 2 and 3 (arms straight) the procedure was the same, except that the experimenter kept the items
 288 concealed between her fingers and thumb instead of in her closed fist, so they were still not visible to the
 289 capuchin, but the presentation was more similar to the way in which food items are normally handed to
 290 the subjects (see video in Online Resource 2; all video captions are in Online Resource 9). In addition,
 291 after the subject had made their selection, the experimenter revealed what item was in the unselected
 292 hand. In session 4 (arms crossed) the procedure was the same as in sessions 2 and 3, except that after
 293 drawing an item from each of the jars, the experimenter crossed her arms over, so that the hand containing
 294 the item from Jar A (a peanut) was next to Jar B (containing 100% pellets) when the subject made their
 295 selection.

296 **Results and discussion**

297 Subjects selected the hand containing the item from Jar A (peanut:pellet ratio of 300:0) in 61.0%
 298 of trials (Figure 3), significantly more than expected by chance (one-sample t-test: $t(18) = 3.713$, $P =$
 299 0.002 , $d = 0.1.750$)¹.

300 * *Figure 3 about here* *

301
 302 A repeated measures ANOVA, with session (1 – 4) as a within-subjects factor found no main effect of
 303 session on performance ($F(3,54) = 1.619$, $P = 0.196$, partial $\eta^2 = 0.183$), suggesting that subjects did not
 304 learn to solve the task over the course of the experiment (performance across trials is shown in Fig. S1a of

¹ All analyses reported in this manuscript were also run using non-parametric tests (Friedman's tests, Wilcoxon signed-rank tests; Mann-Whitney U tests) and produced similar p values in all experiments; see Online Resource 3

PROBABILISTIC INFERENCE IN CAPUCHINS

305 Online Resource 4), and also that the method of presenting the items (closed fists vs. finger/thumb) did
306 not influence performance. Therefore in Experiments 2 – 4 we used the closed fist method, to maximise
307 comparability with the previous ape study (Rakoczy et al. 2014). Trial 1 performance was significantly
308 better than chance, with 16/19 subjects (84.2%) selecting the hand containing the item from Jar A
309 (binomial test: $P < 0.001$), which further supports an absence of learning across trials. Pooling Trial 1
310 performance for each of the four sessions of Experiment 1 also revealed above-chance performance
311 (mean = 3.1 trials correct out of 4; one-sample t-test: $t(18) = 5.144$, $P < 0.001$, $d = 2.425$). Performance
312 did not differ significantly between trials in which the experimenter's arms were straight (61.4% correct)
313 and those in which they were crossed (58.8%; $t(36) = 0.465$, $P = 0.645$, $d = 0.013$), suggesting that
314 subjects were equally able to solve the task regardless of whether the sample was on the same side as the
315 jar it was drawn from, or on the opposite side, and were not simply reaching towards the jar containing
316 the greater quantity of preferred items.

317 Performance in Experiment 1 was poorer than expected overall, given the subjects' strong
318 motivation to obtain peanuts rather than pellets as evidenced by the preference trials. Many subjects
319 exhibited significant side-biases (though there were no 100% side-biased individuals, unlike in
320 Experiments 2 – 4; see Table S1 in Online Resource 1), compared with in the preference trials where
321 none of the subjects were side-biased. Interestingly, recent evidence suggests that making inferences
322 about samples drawn from homogeneous populations can be a non-trivial task, even for 3-year-old
323 children (Giroto et al. 2016). Given that this task did not require subjects to reason about probabilities,
324 this suggests that the sampling procedure, i.e. the experimenter randomly drawing a single item from each
325 population and keeping it hidden in their hand while subjects make their selection poses additional
326 demands (cognitive and/or non-cognitive) that impair performance. This requires knowledge of object
327 permanence (to understand that there were items in the experimenter's hands that were currently out of
328 sight); short-term memory (for which jar each sample was drawn from); and inhibitory control (to prevent
329 impulsive reaching to a side for which the subject has an inherent preference). While apes have not been

PROBABILISTIC INFERENCE IN CAPUCHINS

330 tested in a comparable baseline task, there is some evidence that apes outperform capuchins in tests of
331 object permanence, short-term memory and inhibitory control (Amici et al. 2008; 2010), and we return to
332 this in the General Discussion.

333 **Experiment 2: Inferences from heterogeneous populations to samples**

334 The aim of Experiment 2 was to investigate the ability of capuchins to make an inference about
335 which of two single-item samples drawn from two populations differing in their distributions of preferred
336 to non-preferred items is more likely to consist of a preferred item.

337 **Subjects**

338 Seventeen subjects participated in Experiment 2 (see Table 1). Two subjects did not participate
339 due to a lack of motivation to come into the testing cubicles for sufficient sessions to complete the
340 experiment.

341 **Apparatus and procedure**

342 The jars depicted in Fig. 1b were used. Both jars contained the same total number of items (300)
343 but Jar A contained a 4:1 distribution of peanuts to pellets, and Jar B contained a 1:4 distribution of
344 peanuts to pellets. The samples drawn always consisted of a peanut from Jar A and a pellet from Jar B
345 (the majority item, as in Rakoczy et al. 2014). Items were presented to the subject in closed fists and once
346 they had taken the selected item the alternative item was revealed to them (see video in Online Resource
347 5). There were three sessions of six trials (18 trials in total).

348 **Results and discussion**

349 Of the seventeen subjects that participated in Experiment 2, seven exhibited a 100% side bias
350 (they chose the sample on the same side in all 18 trials), suggesting that their behaviour was independent
351 of the populations in the jars, and thus uninformative with regards to our experimental question.

PROBABILISTIC INFERENCE IN CAPUCHINS

352 Therefore, we excluded these subjects from our analyses; an approach that has previously been used with
353 young children (e.g. Austin et al. 2014), capuchins (e.g. de Waal et al. 2008; Schrauf et al. 2008), and
354 other animal species (e.g. Tauzin et al. 2015) in two-alternative forced choice tasks. We followed this
355 procedure for the remainder of the experiments reported in this paper (results of analyses with 100% side-
356 biased individuals included are available in Online Resource 6).

357 The ten subjects that were not 100% side-biased selected the hand containing the item from Jar A
358 (peanut:pellet ratio of 240:60) in 63.9% of trials (Figure 3); significantly more than expected by chance
359 (one-sample t-test: $t(9) = 3.049$, $P = 0.014$, $d = 2.03$). A repeated measures ANOVA, with session (1 –
360 3) as a within-subjects factor and arms configuration (straight or crossed) as a between-subjects factor
361 found no main effect of session ($F(2,16) = 1.869$, $P = 0.186$, partial $\eta^2 = 0.108$), suggesting that subjects
362 did not learn to solve the task over the course of the experiment (see also Fig. S1b in Online Resource 4).
363 There was also no effect of arms configuration ($F(1,8) = 0.055$, $P = 0.820$, partial $\eta^2 = 0.021$), suggesting
364 that subjects were equally able to solve the task whether the experimenter's arms were straight or crossed.
365 There was no interaction between session and arms configuration ($F(2,16) = 0.486$, $P = 0.624$, partial η^2
366 $= 0.052$).

367 In Trial 1 of the experiment, only 5/10 subjects (50.0%) selected the hand containing the item
368 from Jar A (binomial test: $P = 1.00$). However, pooling Trial 1 performance for each subject across the
369 three sessions of Experiment 2 to increase power revealed performance that was significantly better than
370 chance (mean = 2.1 trials correct out of 3; one-sample t-test: $t(9) = 3.343$, $P = 0.009$, $d = 2.229$).

371 While the results of Experiment 2 suggest that capuchins may be capable of rudimentary
372 probabilistic reasoning, probability and quantity were confounded in this experiment, because more
373 numerous also meant more probable (Denison and Xu 2014). Given that previous work has shown that
374 capuchins are capable of comparing quantities of items and selecting the larger of the two (e.g. Addessi et
375 al. 2008; Evans et al. 2009; VanMarle et al. 2006) it is possible that subjects succeeded by using a

PROBABILISTIC INFERENCE IN CAPUCHINS

376 quantity heuristic such as “select the sample from the jar containing the most peanuts” without
377 considering the proportions in each jar. Therefore, the findings from Experiment 2 replicate results in the
378 animal numerical reasoning literature and extend it, as the monkeys were asked to indicate one of two
379 hidden samples, rather than choose between the distributions themselves, suggesting some understanding
380 of *sampling* and not straightforward numerical comparison. Returning to the question of heuristics, the
381 aim of Experiment 3 was to directly address this possibility.

382 **Experiment 3: Ruling out a choice heuristic based on absolute quantity of preferred items**

383 In this experiment, we pitted absolute quantity of preferred items against probability. If subjects
384 base their selection on the sample from the jar containing the larger absolute quantity of peanuts rather
385 than reasoning about relative proportions, then they should select the sample from Jar B more often than
386 expected by chance.

387 **Subjects**

388 Fifteen subjects participated in Experiment 3 (see Table 1). The other four subjects did not
389 participate due to a lack of motivation to participate in sufficient sessions to complete the experiment.

390 **Apparatus and procedure**

391 The jars depicted in Figure 1c were used. Jar A contained 32 peanuts and 8 pellets (4:1), and Jar
392 B contained 60 peanuts and 240 pellets (1:4). As in Experiment 2, the samples always consisted of a
393 peanut from Jar A and a pellet from Jar B. Items were presented to the subject in closed fists and once
394 they had taken the selected item the alternative item was revealed to them (see video in Online Resource
395 7). There were three sessions of six trials (18 trials in total).

396 **Results and discussion**

397 Of the fifteen subjects that participated in Experiment 3, four exhibited a constant side bias) and
398 so were excluded from our analyses. The eleven subjects that were not 100% side-biased selected the

PROBABILISTIC INFERENCE IN CAPUCHINS

399 hand containing the item from Jar A (peanut:pellet ratio of 32:8) in 67.7% trials correct ($t(10) = 3.791$, P
400 $= 0.004$, $d = 2.40$). A repeated measures ANOVA, with session (1 – 3) as a within-subjects factor and
401 arms configuration (straight or crossed) as a between-subjects factor found no main effect of session
402 ($F(2,18) = 0.10$, $P = 0.990$, partial $\eta^2 = 0.001$), suggesting that subjects did not learn to solve the task over
403 the course of the experiment (see also Fig. S1c in Online Resource 4). There was also no effect of arms
404 configuration ($F(1,9) = 0.003$, $P = 0.955$, partial $\eta^2 = 0.000$), suggesting that subjects were equally able to
405 solve the task whether the experimenter's arms were straight or crossed. There was no interaction
406 between session and arms configuration ($F(2,9) = 1.573$, $P = 0.241$, partial $\eta^2 = 0.149$).

407 In Trial 1 of Experiment 3, 7/11 subjects (63.6%) selected the hand containing the item from Jar
408 A (binomial test: $P = 0.549$). Pooling Trial 1 performance for each subject across the three sessions of
409 Experiment 3 to increase power revealed performance that was significantly better than chance (mean =
410 2.0 trials correct out of 3; one-sample t-test: $t(10) = 2.622$, $P = 0.026$, $d = 1.658$).

411 The results of Experiment 3 further support the idea that capuchins are capable of rudimentary
412 probabilistic reasoning, as they were able to make accurate inferences about samples drawn from
413 populations that were not based on the absolute quantity of preferred items, as has been demonstrated
414 with infants (Denison and Xu 2014) and apes (Rakoczy et al. 2014). However, there are two additional
415 heuristics that capuchins could still potentially have used to make decisions in Experiments 2 and 3, and
416 which infants and apes could have used in previous studies, which are impossible to tease apart from
417 probabilistic inference given the distributions used in those experiments. First, the possibility remains that
418 capuchins could have succeeded in both experiments by avoiding the sample from the jar containing the
419 larger absolute quantity of non-preferred items; e.g. by using a heuristic such as “select the sample from
420 the jar containing the fewest pellets”; a possibility previous work with other species does not address,
421 though Rakoczy and colleagues (2014) do discuss it. This alternative would allow them to avoid
422 comparing the *ratio* of peanuts to pellets in Jar A to the *ratio* of peanuts to pellets in Jar B, and allow
423 them instead to compare the absolute quantities of pellets across jars. Alternatively, subjects could have

PROBABILISTIC INFERENCE IN CAPUCHINS

424 used a different heuristic, one slightly more complex than the avoidance strategy but still a shortcut to
425 engaging in true comparison of proportions. In both Experiments 2 and 3, capuchins were faced with a
426 decision between a sample drawn from a jar containing a larger quantity of peanuts than pellets versus a
427 sample from a jar containing a larger quantity of pellets than peanuts. They could avoid comparing the
428 ratios in each jar to one another by simply marking any jar that has a larger number of peanuts than pellets
429 a “good” jar, and any jar that has a larger number of pellets than peanuts a “bad” jar. In this case,
430 comparison of ratios across jars is unnecessary, as subjects can simply select the sample drawn from the
431 good jar (or avoid the sample from the bad jar) rather than compare ratios (Denison and Xu 2014). We
432 address both of these potential heuristics in Experiment 4. Jar A contained 100 peanuts and 200 pellets
433 and Jar B contained 22 peanuts and 200 pellets. This addresses the first heuristic based on avoiding
434 pellets, as the jars have equal absolute quantities of pellets. Thus if capuchins use absolute quantity
435 estimations to avoid pellets, they will perform at chance. It addresses the second heuristic because, if a
436 subject were simply labelling jars as “good” or “bad”, he would have to label both of these jars as “bad”,
437 as they both contain more pellets than peanuts, and they would not know which sample to select, again
438 performing at chance. If they instead can compare the ratios of peanuts to pellets, then they should be
439 more likely to select the sample from Jar A. Experiment 4 thus represents a particularly challenging case
440 that no species, including human infants, has yet been shown to solve.

441 **Experiment 4: Ruling out a choice heuristic based on avoiding the larger absolute quantity**
442 **of non-preferred items, or labelling jars as “good” and “bad”**

443 Experiment 4 was a novel experimental condition that infants and apes have not previously been
444 tested on, which aimed to investigate, for the first time, whether individuals might potentially be using an
445 alternative heuristic (as opposed to choosing on the basis of the greater quantity of preferred items, which
446 has been ruled out by Experiment 3) when solving this type of task. In this experiment we kept the
447 absolute quantity of non-preferred items the same in both jars, and also in the majority, so both jars would
448 be “bad” jars. Therefore, if subjects were basing their selection on avoiding the jar containing the greater

PROBABILISTIC INFERENCE IN CAPUCHINS

449 absolute quantity of non-preferred items, or were simply labelling jars as “bad” and avoiding them, they
450 would be expected to perform at chance-level (50% of trials correct).

451 Subjects

452 Sixteen subjects participated in Experiment 4 (see Table 1). The remaining three subjects did not
453 participate due to a lack of motivation to participate in sufficient sessions to complete the experiment.

454 Apparatus and procedure

455 The jars depicted in Figure 1d were used. Jar A contained 100 peanuts and 200 pellets, and Jar B
456 contained 22 peanuts and 200 pellets. Unlike in Experiments 1 – 3 where Jar A always contained a greater
457 quantity of peanuts than pellets whereas the reverse was true for Jar B, in Experiment 4 both jars
458 contained a greater quantity of pellets than peanuts. Therefore, we chose to manipulate the sample drawn
459 from Jar A so that unlike in Experiments 1 – 3 it did not consist of a peanut in every trial; instead a peanut
460 was drawn from Jar A in 2/6 trials, and a pellet in the remaining 4/6 trials (to match the probability of the
461 population). The order in which the different items were drawn out of Jar A for the different sessions was
462 the same for each monkey and as follows: session 1: pellet, peanut, pellet, pellet, peanut, pellet; session 2:
463 peanut, pellet, pellet, peanut, pellet, pellet; session 3: pellet, pellet, peanut, pellet, pellet, peanut. A pellet
464 (the majority item) was always drawn out of Jar B. Items were presented to the subject in closed fists and
465 once they had taken the selected item the alternative item was revealed to them (see video in Online
466 Resource 8). There were three sessions of six trials (18 trials in total).

467 Results and discussion

468 Of the sixteen subjects that participated in Experiment 4, seven exhibited a constant side bias and
469 so were excluded from our analyses. The nine subjects that were not 100% side-biased selected the hand
470 containing the item from Jar A (peanut:pellet ratio of 100:200) in 58.0% of trials (Figure 3), and while
471 performance was in the same direction as the other experiments, it was only marginally significant (one-

PROBABILISTIC INFERENCE IN CAPUCHINS

472 sample t-test: $t(8) = 2.163$, $P = 0.063$, $d = 1.53$). A repeated measures ANOVA (corrected using
473 Greenhouse-Geisser estimates of sphericity), with session (1 – 3) as a within-subjects factor and arms
474 configuration (straight or crossed) as a between-subjects factor found no main effect of session
475 ($F(1.123,7.862) = 0.758$, $P = 0.425$, partial $\eta^2 = 0.098$), suggesting that subjects did not learn to solve the
476 task over the course of the experiment (see also Fig. S1d in Online Resource 4). There was also no effect
477 of arms configuration ($F(1,7) = 0.012$, $P = 0.916$, partial $\eta^2 = 0.002$), suggesting that subjects were
478 equally able to solve the task whether the experimenter's arms were straight or crossed. There was no
479 interaction between session and arms configuration ($F(1.123,7.862) = 0.408$, $P = 0.645$, partial $\eta^2 =$
480 0.055).

481 In Trial 1 of Experiment 4, 3/9 subjects (33.3%) selected the hand containing the item from Jar A
482 (binomial test: $P = 0.508$). Pooling Trial 1 performance for each subject across the three sessions of
483 Experiment 4 to increase power revealed performance that did not differ significantly from chance (mean
484 = 1.67 trials correct out of 3; one-sample t-test: $t(8) = 0.577$, $P = 0.580$, $d = 0.408$).

485 Although capuchins' performance in Experiment 4 was only marginally above chance and Trial 1
486 performance did not differ from chance, additional factors unrelated to probabilistic reasoning may have
487 contributed to making the task presented in Experiment 4 more challenging than Experiments 1 – 3. First,
488 the populations in Jar A and B were more difficult to discriminate visually than in the other experiments
489 since both contained a majority of pellets (see Online Resource 10). Second, the reward schedule
490 implemented differed from that used in the other experiments (as described in the Apparatus and
491 Procedure section for Experiment 4). In Experiments 1 – 3 the sample always consisted of the more
492 probable item from each jar (as in Rakoczy et al. 2014), which in each case was a peanut from Jar A and a
493 pellet from Jar B. Therefore subjects were always rewarded for selecting the hand containing the sample
494 from the "correct" jar. In Experiment 4 however, because the most probable item from each jar would
495 have been a pellet on every trial, we manipulated the sample drawn from Jar A to match the probability of
496 the population, such that it consisted of a peanut in 2/6 trials. This reward schedule would be less likely to

PROBABILISTIC INFERENCE IN CAPUCHINS

497 result in reinforcement-based learning, and may have led to reduced motivation. However, Trial 1
498 performance in Experiment 4 (33.3% correct) was also lower in this experiment than overall performance,
499 and lower than Trial 1 performance in Experiments 1 – 3, which cannot be explained by the different
500 reinforcement schedule.

501 **General discussion**

502 The results of these experiments suggest that some capuchin monkeys, like human infants
503 (Denison and Xu 2010; 2014) and great apes (Rakoczy et al. 2014), are capable of making probabilistic
504 inferences from populations to samples, and success was not due to learning across trials. In particular,
505 Experiment 3 ruled out the possibility that capuchins succeeded by using a heuristic based on comparing
506 absolute quantities of preferred items in the two populations², though relatively poor performance in
507 Experiment 4 suggests that capuchins (and possibly infants and apes) may rely on quantity-based
508 heuristics in certain situations. Our experiments show that at minimum capuchins do not solve these tasks
509 by using one simple heuristic that even school-aged children have been shown to rely on in some more
510 explicit probabilistic inference tasks (Falk et al., 2012): selecting the item from the population containing
511 the greatest absolute quantity of preferred items. The possibility remains that individuals of any of the
512 taxa tested to date could be flexibly using a combination of different heuristics in different tasks (e.g. in
513 our study “select sample from population with greatest absolute quantity of peanuts” in Experiments 2
514 and 4, and “avoid sample from population with greatest absolute quantity of pellets” in Experiment 3).
515 However, we believe that probabilistic inference is a more parsimonious explanation for our data. Taken
516 together, our results provide some evidence to suggest that the capacity for rudimentary intuitive statistics
517 may be evolutionarily ancient, given that humans and capuchins shared a common ancestor over 30

² It should be noted that it is not possible to say *how* capuchins were estimating proportions, i.e. whether they computed probabilities over numerical representations or continuous quantities, and this question has not yet been examined in either infants or apes. While this is an interesting avenue for future research, computing proportions is about considering relative amounts, regardless of format.

PROBABILISTIC INFERENCE IN CAPUCHINS

518 million years ago (Fragaszy, 2004; though it is also possible that this capacity evolved convergently in
519 capuchins and apes, Reader et al. 2011).

520 Despite some methodological differences between species (e.g. number of subjects, number of
521 trials, exact quantities of items in populations) it is possible to draw meaningful comparisons between the
522 results of the current capuchin study and previous data from infants and apes. Infants, apes and capuchins
523 were all presented with a task where the total number of items in the two populations was the same, but
524 the proportions of preferred to non-preferred items were reversed (4:1 vs. 1:4 for all three species; total
525 number of items in the populations varied between species). Infants were only presented with a single
526 trial (Denison and Xu 2010); therefore their performance can be compared with Trial 1 performance for
527 apes and capuchins. Because side-bias data were not available for the previous ape study, here we discuss
528 our own data with all capuchins included, to facilitate valid comparison. Twenty-five out of 32 infants
529 (78%) succeeded in their single trial (Denison and Xu 2010), compared with correct Trial 1 performance
530 by 20/28 apes (71%; Rakoczy et al. 2014: Experiment 1). In our study (Experiment 2), 10/17 capuchins
531 (59%) chose correctly in Trial 1, with 7 of those individuals subsequently exhibiting a 100% side-bias (5
532 who chose correctly and 2 incorrectly in Trial 1). Pooling data for all trials, both capuchins and great apes
533 performed above chance-level, though capuchins succeeded in fewer trials (58% correct with completely
534 side-biased individuals' data included; Experiment 2 of this study), than apes (71% correct; Rakoczy et al.
535 2014: Experiment 1).

536 All three species were also presented with a task in which absolute quantity was pitted against
537 probability (as in Experiment 3 of this study); such that the population that was more likely to produce a
538 preferred-item sample contained the smaller absolute quantity of preferred items (though again total
539 numbers of items in the populations varied between species). Nineteen out of 24 infants (79%) succeeded
540 in their single trial (Denison and Xu 2014: Experiment 2), compared with correct Trial 1 performance by
541 20/26 apes (77%, Rakoczy et al. 2014). In our Experiment 3, 9/15 capuchins chose correctly in Trial 1,
542 with 4 of these individuals (2 that chose correctly and 2 incorrectly) subsequently exhibiting a 100% side-

PROBABILISTIC INFERENCE IN CAPUCHINS

543 bias. Across all trials, apes succeeded in 63% (Rakoczy et al. 2014: Experiment 6), which was the same as
544 capuchins' success rate (63% of trials correct with completely side-biased individuals' data included;
545 Experiment 3 of this study), providing evidence of an ability to reason about *relative* frequencies of
546 preferred and non-preferred items within populations and to draw inferences about random single-item
547 samples drawn from these populations.

548 In our Experiment 4, which was a novel condition unexamined in previous work done with
549 infants and apes, capuchins were presented with a task in which the total number of non-preferred items
550 was held constant in the two populations, and also outnumbered the preferred items in both jars (i.e. both
551 jars were unlikely to yield a preferred item, but one was more unlikely than the other). This meant that
552 subjects could not succeed by avoiding the sample drawn from the population containing the greater
553 quantity of non-preferred items, or by marking one Jar As “bad” and one Jar As “good”. Capuchins’
554 performance was marginally different from chance across all trials (58% of trials correct overall). Trial 1
555 performance did not differ from chance (3/9 subjects, 33%, correct), even when Trial 1 of each session
556 was pooled to increase power (56% of first trials correct). Apes were not tested in a task where the
557 quantity of non-preferred items was equal in the two populations (Rakoczy et al. 2014) and infant
558 performance was only marginally significant in an analogous task in which both populations were likely
559 to yield a preferred object but one was more likely (Denison and Xu 2014: Experiment 4), which suggests
560 that there may be something more difficult about this task. One possibility (in addition to the different
561 reinforcement schedule mentioned in the Experiment 4 Results and Discussion section) is that the ratio
562 between ratios (defined as the ratio of preferred to non-preferred items in the favourable population,
563 divided by the ratio of preferred to non-preferred items in the unfavourable population; Drucker et al.
564 2016) of the populations in Experiment 4 ($(100/200) / (22/200) = 4.55$) was lower than in Experiment 2
565 ($(240/60) / (60/240) = 16$) and Experiment 3 ($(32/8) / (60/240) = 16$). Drucker et al. (2016) found that
566 macaques were better able to select the “favourable” of two arrays (greater ratio of positive to negative
567 stimuli) on a touchscreen when the ratio between ratios was higher. It also leaves open the possibility that

PROBABILISTIC INFERENCE IN CAPUCHINS

568 at least some subjects may have used a strategy that involved avoiding the population containing the
569 greatest absolute quantity of non-preferred items (not possible in Experiment 4 as both populations
570 contain the same number of non-preferred items), or by marking each Jar As “good” or “bad” (both jars
571 would be “bad” in this task).

572 A critic could argue that capuchins solved the tasks presented in our study due to a “Clever Hans”
573 effect; that is, by using subtle behavioural cues from the human experimenter. However, we think this is
574 unlikely for the following reasons. First, Rakoczy et al. (2014) demonstrated that apes still solved this
575 type of task when Clever Hans effects were controlled for in an intricately designed experiment involving
576 two experimenters and special equipment (which we were unfortunately unable to replicate due to testing
577 setup constraints), and their performance did not differ from the original experiment without the Clever
578 Hans controls. Second, non-human primates are notoriously poor at understanding even deliberate human
579 communicative cues such as pointing (e.g. Bräuer et al. 2006; but see e.g. Hopkins et al. 2013 for
580 evidence that chimpanzees can utilise human pointing as a cue in a modified object-choice task), and
581 capuchins specifically were found to be unable to use experimenter gaze direction to locate a food reward
582 hidden under one of two objects, even after receiving 510 trials (Anderson et al. 1995). Finally, as
583 described in the Methods section the experimenter was aware of the potential for unintentional cueing and
584 implemented several measures to control for this possibility.

585 It could also be argued that capuchins solved the tasks by using olfactory cues from the items
586 concealed in the experimenter’s hands; however we also think this is unlikely. Capuchins rely on visual
587 information more than olfactory cues to locate food (Fragaszy et al. 2004), and free-ranging capuchins did
588 not succeed in using olfactory cues to locate food concealed in containers (Bolen and Green 1997). Our
589 experimental set-up also made it difficult for capuchins to exploit olfactory cues. The cubicle doors were
590 polycarbonate windows with small arm holes (as opposed to more open wire mesh), and at the time of
591 choice the samples were held at such a distance that the capuchins had to fully extend an arm out of the
592 window to reach one of the experimenter’s hands (see videos in Online Resources 2, 5, 7 and 8) meaning

PROBABILISTIC INFERENCE IN CAPUCHINS

593 that it was not possible to sniff the samples directly, making the detection of odour cues was unlikely.
594 Additionally, the latex gloves worn by the experimenter at all times had a strong odour, and all testing
595 sessions started with the test individual being handed both sunflower seeds and raisins, adding further
596 scents to the experimenter's gloved hands.

597 As mentioned earlier (see Experiment 1 Results and Discussion), capuchins performed more
598 poorly than we expected in the baseline condition, which did not require subjects to reason
599 probabilistically. Any factors limiting performance in Experiment 1 would also apply to Experiments 2 –
600 4 where subjects were additionally required to reason about proportions, and therefore could also have
601 been responsible for limiting performance in these test conditions. Is there any evidence that apes perform
602 better in tasks designed to test object permanence, short-term memory or inhibitory control that might
603 explain capuchins' relatively lower success rate in some of the sampling tasks?

604 Amici and colleagues (2008, 2010) compared the performance of several ape and monkey species
605 in a battery of physical cognition and inhibitory control tasks. Chimpanzees and bonobos outperformed
606 capuchins in a short-term memory task (though capuchins still performed above chance-level), and
607 capuchins were outperformed by chimpanzees, bonobos and gorillas in a single invisible displacement
608 task (Amici et al. 2010). Similarly, capuchins performed significantly worse than chimpanzees and
609 bonobos in a series of inhibitory control tasks (Amici et al. 2008; but see MacLean et al. 2014 for
610 evidence of capuchins performing comparably to great apes in two inhibitory control tasks).

611 Interestingly, Girotto et al. (2016) recently presented 3-year-old children with a task comparable
612 to our Experiment 1, where one population consisted of 100% of one type of item, and the second
613 population consisted of 100% of another type of item. Children were presented with a single trial, and
614 unlike for our capuchins, the samples were never crossed over (so the task was arguably more
615 straightforward). In one of these tasks (Study 2, Task A; the one that was most similar to our Experiment
616 1), only 33 out of 48 3-year-olds selected the sample drawn from the favourable population (69% correct),

PROBABILISTIC INFERENCE IN CAPUCHINS

617 compared with capuchins' Trial 1 performance of 16/19 (84%) correct. This finding suggests that making
618 inferences from homogeneous populations to samples can be a non-trivial task, even for 3-year-old
619 children, and the authors posit that this may be due to inhibitory control limitations resulting in a working
620 memory overload (Giroto et al. 2016).

621 The presence of significant side-biases throughout this study, and evidence from comparative
622 studies that apes outperform capuchins in tasks that rely on abilities related to our choice-based dependent
623 variable (e.g. object permanence, short-term memory and inhibitory control; Amici et al. 2008; 2010)
624 suggest that it was not having to reason about probabilities that was more challenging for the capuchins
625 than the infants and apes previously tested using this paradigm. This is further supported by the fact that
626 capuchins' performance in Experiments 2 and 3 of this study did not differ from their performance in
627 Experiment 1 (baseline condition), which did not involve probabilistic reasoning. One way to investigate
628 this further would be to test capuchins on looking-time versions of our experiments. If capuchins
629 performed better in this version than in our current action-based version (i.e. they reliably looked longer
630 at unlikely samples) then this would bolster the claim that it is these other aspects of the task, not
631 reasoning about probabilities, that limits capuchins' performance.

632 In conclusion, we found evidence that at least some capuchins, like human infants and apes, were
633 able to make inferences about single-item samples randomly drawn from heterogeneous populations
634 (Experiment 2), and this was achieved by reasoning about relative as opposed to absolute frequencies of
635 preferred and non-preferred items within populations (Experiment 3). This is the first evidence for
636 intuitive probabilistic inference in a monkey species, suggesting that the ability to reason about
637 probabilities may be evolutionarily ancient. However, given that sophisticated cognitive abilities may
638 have evolved convergently in capuchins and great apes (Reader et al. 2011), additional primate species
639 would need to be tested to establish just how widespread the capacity for probabilistic inference is. As
640 performance was relatively poor in Experiment 4 – our novel experimental condition that goes beyond the
641 work previously done with either apes or infants – further research is required to establish whether some

PROBABILISTIC INFERENCE IN CAPUCHINS

642 capuchins might use strategies that involve avoiding non-preferred items or marking the populations as
643 “good” and “bad”; and whether the same might be true for apes and/or infants. Given the broad ecological
644 relevance of reasoning about proportions, future research should also aim to investigate whether
645 probabilistic inference is an ability that is also shared with non-primate species.

646 **Acknowledgements**

647 We would like to thank the RZSS Living Links staff who gave permission to conduct research and
648 provided invaluable support and assistance. Thanks to Keith Haynes at the University of St Andrews for
649 constructing the cubicle window. Thanks also to Natalia Robert-Nunez for secondary coding of data.

650 **Compliance with Ethical Standards**

651 This study was funded by an Economic and Social Research Council Future Research Leaders grant
652 (ES/K009540/1) to DB. The authors declare that they have no conflict of interest. All procedures were in
653 accordance with UK law and the ASAB Guidelines for the Treatment of Animals in Behavioural
654 Research and Teaching. The study was approved by the University of St Andrews School of Psychology
655 & Neuroscience Ethics Committee and the Research Committee at Living Links.

656 **References**

657 Addressi E, Crescimbene L, Visalberghi E (2008) Food and token quantity discrimination in capuchin
658 monkeys (*Cebus apella*). *Anim Cogn* 11:275–82.

659 Amici F, Aureli F, Call J (2008) Fission-fusion dynamics, behavioural flexibility, and inhibitory control
660 in primates. *Curr Biol* 18:1415–1419.

661 Amici F, Aureli F, Call J (2010) Monkeys and apes: Are their cognitive skills really so different? *Am J*
662 *Phys Anthropol* 143:188–197.

PROBABILISTIC INFERENCE IN CAPUCHINS

- 663 Anderson JR, Sallaberry P, Barbier H (1995) Use of experimenter-given cues during object-choice tasks
664 by capuchin monkeys. *Anim Behav* 49:201–208.
- 665 Austin K, Theakston A, Lieven E, Tomasello M (2014) Young children’s understanding of denial. *Devel*
666 *Psychol*, 50(8), 2061-2070.
- 667 Beran MJ, Decker S, Schwartz A, Schultz N (2011) Monkeys (*Macaca mulatta* and *Cebus apella*) and
668 human adults and children (*Homo sapiens*) compare subsets of moving stimuli based on
669 numerosity. *Front Psychol* 2:61.
- 670 Bolen RH, Green SM (1997). Use of olfactory cues in foraging by owl monkeys (*Aotus nancymai*) and
671 capuchin monkeys (*Cebus apella*). *J Comp Psychol* 111:152–158.
- 672 Bräuer J, Kaminski J, Riedel J, Call J, Tomasello M (2006) Making inferences about the location of
673 hidden food: social dog, causal ape. *J Comp Psychol*, 120:38–47.
- 674 Brosnan SF, de Waal FBM (2003) Monkeys reject unequal pay. *Nature* 425:297-299.
- 675 Brosnan SF, Schiff HC, de Waal FBM (2005) Tolerance for inequity may increase with social closeness
676 in chimpanzees. *Proc Roy Soc B: Biol Sci* 272 :253-258
- 677 Bryant P, Nunes T (2012) Children’s understanding of probability: A literature review (full report). A
678 report to the Nuffield Foundation. Retrieved on 05/20/2015 from:
679 http://www.nuffieldfoundation.org/sites/default/files/files/Nuffield_CuP_FULL_REPORTv_FIN
680 [AL.pdf](#).
- 681 Cronin KA, Snowdon CT (2008) The effects of unequal reward distributions on cooperative problem
682 solving by cotton-top tamarins (*Saguinus oedipus*). *Anim Behav* 75: 245-257.
- 683 Denison S, Xu F (2010) Twelve- to 14-month-old infants can predict single-event probability with large
684 set sizes. *Dev Sci* 13:798–803.

PROBABILISTIC INFERENCE IN CAPUCHINS

- 685 Denison S, Xu F (2014) The origins of probabilistic inference in human infants. *Cognition* 130:335–347.
- 686 de Waal FBM, Leimgruber K, Greenberg AR (2008) Giving is self-rewarding for monkeys. *Proc Natl Acad Sci* 105: 13685–13689.
- 687
- 688 Drucker CB, Rossa MA, Brannon EM (2016) Comparison of discrete ratios by rhesus macaques (*Macaca mulatta*). *Anim Cogn* 19:57–89.
- 689
- 690 Evans TA, Beran MJ, Harris EH, Rice DF (2009) Quantity judgments of sequentially presented food
- 691 items by capuchin monkeys (*Cebus apella*). *Anim Cogn* 12:97–105.
- 692 Falk R, Yudilevich-Assouline P, Elstein A (2012) Children’s concept of probability as inferred from their
- 693 binary choices – revisited. *Educational Studies in Mathematics* 81:207–233.
- 694 Fontanari L, Gonzalez M, Vallortigara G, Girotto V (2014) Probabilistic cognition in two indigenous
- 695 Mayan groups. *Proc Natl Acad Sci* 111: 17075–17080.
- 696 Frigaszy DM, Visalberghi E, Fedigan LM (2004) *The Complete Capuchin: The Biology of the Genus*
- 697 *Cebus*. Cambridge: Cambridge University Press.
- 698 Girotto V, Fontanari L, Gonzalez M, Vallortigara G, Blaye A (2016) Young children do not succeed in
- 699 choice tasks that imply evaluating chances. *Cognition* 152:32–39.
- 700 Hanus D, Call J (2014) When maths trumps logic : probabilistic judgements in chimpanzees. *Biol Lett*
- 701 10:20140892.
- 702 Hopkins WD, Russell JL, McIntyre J, Leavens DL (2013) Are chimpanzees really so poor at
- 703 understanding imperative pointing? Some new data and an alternative view of canine and ape
- 704 social cognition. *PLoS One* 8(11), 379338.
- 705 Jordan KE, Brannon EM (2006) Weber’s Law influences numerical representations in rhesus macaques
- 706 (*Macaca mulatta*). *Anim Cogn* 9:159–172.

PROBABILISTIC INFERENCE IN CAPUCHINS

- 707 Judge PG, Evans TA, Vyas DK (2005) Ordinal representation of numeric quantities by brown capuchin
708 monkeys (*Cebus apella*). J Exp Psychol: Anim Behav Proc, 31:79–94.
- 709 Leonardi R, Buchanan-Smith HM, Dufour V, MacDonald C, Whiten A (2010). Living together:
710 behaviour and welfare in single and mixed species groups of capuchin (*Cebus apella*) and squirrel
711 monkeys (*Saimiri sciureus*). Am J Primatol 72:33–47.
- 712 MacLean EL, Hare B, Nunn CL et al. (2014) The evolution of self-control. Proc Nat Acad
713 Sci 111:E2140–E2148.
- 714 Maginnity ME, Grace RC (2014) Visual perspective taking by dogs (*Canis familiaris*) in a Guess-
715 Knower task: evidence for a canine theory of mind? Anim Cogn 17:1375–92.
- 716 Piaget J, Inhelder B (1975) The Origin of the Idea of Chance in Children. Routledge and Kegan Paul.
717 Translation of original work (1951).
- 718 Rakoczy H, Clüver A, Saucke L, Stoffregen N, Gräbener A, Migura J, Call J (2014) Apes are intuitive
719 statisticians. Cognition 131:60–68.
- 720 Range F, Horn L, Viranyi Zs, Huber L (2009) Effort and reward: Inequity aversion in domestic dogs?
721 Proc Nat Acad Sci 106:340-345.
- 722 Reader SM, Hager Y, Laland KN (2011) The evolution of primate general and cultural intelligence. Phil
723 Trans R Soc B 366:1017-1027.
- 724 Reznikova Z, Ryabko B (2011) Numerical competence in animals, with an insight from
725 ants. Behaviour 148:405–434.
- 726 Rugani R, McCrink K., de Hevia MD, Vallortigara G, Regolin, L (2016). Ratio abstraction over discrete
727 magnitudes by newly hatched domestic chicks (*Gallus gallus*). Sci Rep 6: 30114.
- 728

PROBABILISTIC INFERENCE IN CAPUCHINS

- 729 Rugani R, Vallortigara G, Regolin L (2015) The use of proportion by young domestic chicks (*Gallus*
730 *gallus*). *Anim Cogn* 18:605–16.
- 731 Schmidjell T, Range F, Huber L, Virányi Z (2012) Do owners have a clever hans effect on dogs? Results
732 of a pointing study. *Front Psychol* 3:558.
- 733 Schmitt V, Fischer J (2011) Representational format determines numerical competence in monkeys.
734 *Nature Communications* 2:257.
- 735 Schrauf C, Huber L, Visalberghi E (2008) Do capuchin monkeys use weight to select hammer tools?
736 *Anim Cogn* 11:413–422.
- 737 Teglas E, Girotto V, Gonzalez M, Bonatti LL (2007). Intuitions of probabilities shape expectations about
738 the future at 12 months and beyond. *Proc Nat Acad Sci* 104:19156–19159.
- 739 Teglas E, Vul E, Girotto V, Gonzalez M, Tenenbaum JB, Bonatti LL (2011) Pure reasoning in 12-month-
740 old infants as probabilistic inference. *Science* 332:1054–1058.
- 741 Vallortigara G (2014) Foundations of Number and Space Representations in Non-Human Species. In
742 "Evolutionary Origins and Early Development of Number Processing", pp. 35-66 (Eds., D.C.
743 Geary, D.B. Bearch, K. Mann Koepke), Elsevier, New York.
- 744 VanMarle K, Aw J, McCrink K, Santos LR (2006) How Capuchin Monkeys (*Cebus apella*) Quantify
745 Objects and Substances. *J Comp Psychol* 120 :416–426.
- 746 Wilson ML, Hauser MD, Wrangham RW (2001) Does participation in intergroup conflict depend on
747 numerical assessment, range location, or rank for wild chimpanzees? *Anim Behav* 6:1203–1216.
- 748 Xu F, Garcia V (2008) Intuitive statistics by 8-month-old infants. *Proc Nat Acad Sci* 105:5012–5015.
- 749

PROBABILISTIC INFERENCE IN CAPUCHINS

750 **Table 1** Details of capuchins that participated in this study. All individuals were born in captivity and mother-
 751 reared, except for Kato who was wild-born and hand-reared. Group refers to the geographical location of the
 752 enclosures at the zoo and age is in years

Name	Group	Sex	Age	Experiment participation
Alba	West	F	2	1,2,3,4
Anita	East	F	16	1,2,3,4
Carlos	East	M	8	1,2,3,4
Chico	East	M	4	1,2,3,4
Diego	West	M	11	1,2,3,4
Figo	West	M	7	1,2,3,4
Flojo	East	M	2	1,2,3,4
Inti	West	M	4	1,2,3,4
Junon	East	F	13	1,2,3,4
Kato	East	M	8	1,2,3,4
Luna	West	F	2	1,2,3,4
Manuel	East	M	8	1,2
Pedra	West	F	5	1
Reuben	East	M	3	1,2,3,4
Rufo	West	M	4	1,2,3,4
Sylvie	West	F	10	1,2,4
Toka	West	M	9	1
Torres	West	M	2	1,2,3,4
Ximo	West	M	3	1,2,3,4

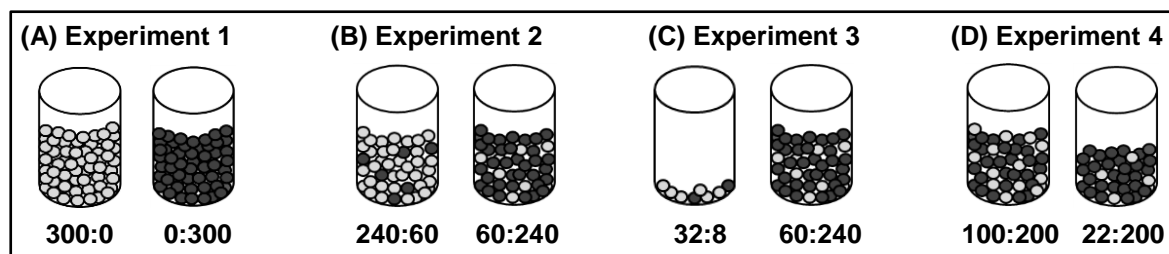
753

PROBABILISTIC INFERENCE IN CAPUCHINS

754

755

756



757

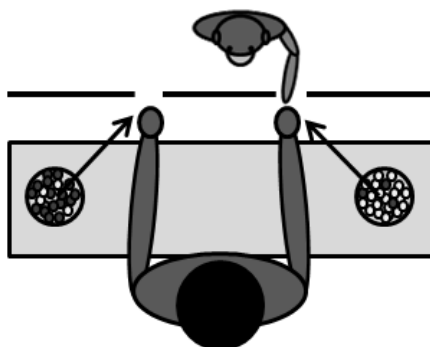
758 **Figure 1** Schematic representations of the distributions of populations in Jar A (left in each pair) and Jar B (right in
 759 each pair) for Experiments 1 – 4 (jar placement was counterbalanced on the left and right in all experiments). Light
 760 grey circles represent peanuts (preferred food item) and dark grey circles represent monkey pellets (non-preferred
 761 food item). Ratios underneath the jars represent the peanut:pellet ratio in that jar. All jars were transparent so the
 populations were continuously visible to the monkeys

762

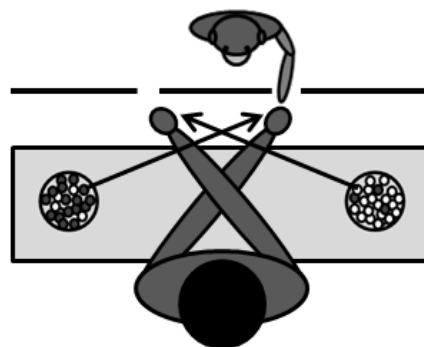
PROBABILISTIC INFERENCE IN CAPUCHINS

763

(a) arms straight



(b) arms crossed

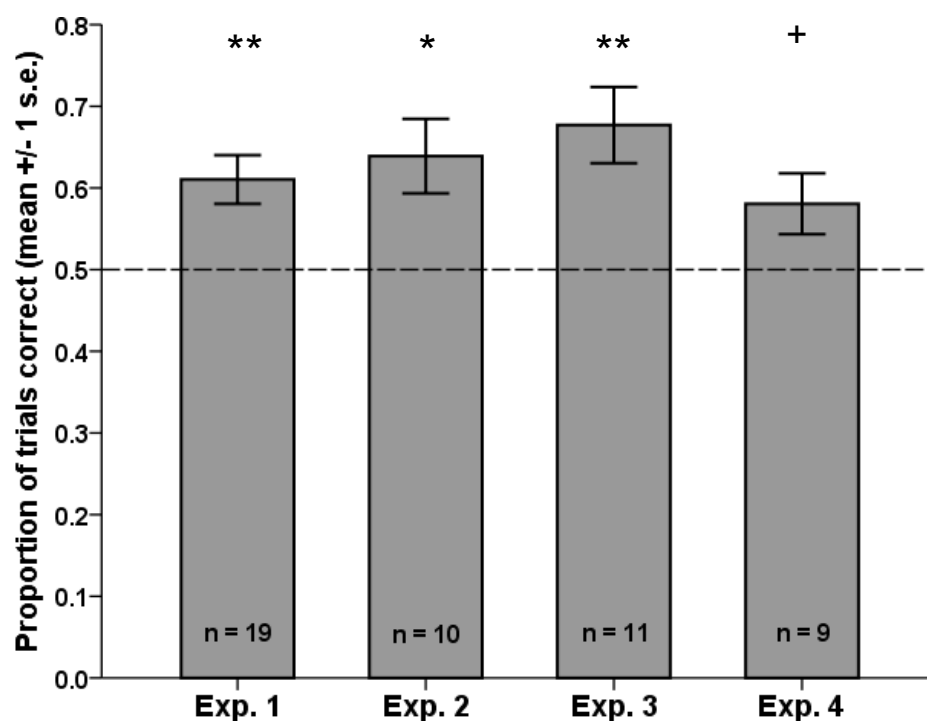


764

765 **Figure 2** Schematic representation of the experimental setup and general procedure. Subjects participated
 766 individually in a test cubicle (see Leonardi et al. 2010 for full details of the cubicle set up) with a custom-made
 767 Plexiglas window. At the start of each trial the experimenter simultaneously shook both jars whilst looking at the
 768 subject to draw their attention. She then randomly drew a single item from each jar simultaneously, and kept them
 769 hidden from the subject in her closed fists. The experimenter then extended her arms to present her closed fists
 770 containing the concealed items at the two holes in the cubicle window, either keeping her arms straight (a) or
 771 crossing them over (b). In Experiment 1, the experimenter's arms were straight for the first three sessions of trials
 772 and crossed for the fourth session. For each of Experiments 2 – 4 subjects were pseudorandomly assigned to either
 773 arms straight or arms crossed presentation, with the constraint that across these three experiments subjects either had
 774 arms straight in 2/3 experiments and arms crossed in 1/3, or vice-versa

775

776



777
 778 **Figure 3** Mean proportion of trials (± 1 standard error) in which subjects selected the hand containing the item from
 779 Jar A in Experiments 1 – 4. Experiment 1 had 24 trials and Experiments 2 – 4 each had 18 trials. All subjects
 780 completed Experiment 1 first; the order in which subjects subsequently completed Experiments 2 – 4 was
 781 randomised. ** indicates $P < 0.01$, * indicates $P < 0.05$, and + indicates $P < 0.07$ in a one-sample t-test. Dashed line
 782 indicates chance-level performance (half of the trials correct). This graph excludes individuals with a 100% side bias
 783 in a given experiment