Rational After All: Changes in Probability Matching Behaviour Across Time in Humans and Monkeys

Carmen Saldana (carmen.saldanagascon@uzh.ch)
Centre for Language Evolution, University of Edinburgh, 3 Charles Street, EH8 9AD, UK
Department of Comparative Linguistics, University of Zurich, Plattenstrasse 54, 8032, Switzerland
Center for the Interdisciplinary Study of Language Evolution , University of Zurich, Plattenstrasse 54, 8032, Switzerland

Nicolas Claidière (nicolas.claidiere@univ-amu.fr) and Joël Fagot (joel.fagot@univ-amu.fr)
Aix Marseille Université, CNRS, LPC UMR 7290, 13331, France
Brain and Language Research Institute, Aix-Marseille University, France

Kenny Smith (kenny.smith@ed.ac.uk)
Centre for Language Evolution, University of Edinburgh, 3 Charles Street, EH8 9AD,UK

Abstract

Probability matching—where subjects given probabilistic input respond in a way that is proportional to those input probabilities—has long been thought to be characteristic of primate performance in probability learning tasks in a variety of contexts, from decision making to the learning of linguistic variation in humans. However, such behaviour is puzzling because it is not optimal in a decision theoretic sense; the optimal strategy is to always select the alternative with the highest positive-outcome probability, known as maximising (in decision making) or regularising (in linguistic tasks). While the tendency to probability match seems to depend somewhat on the participants and the task (i.e., infants are less likely to probability match than adults, monkeys probability match less than humans, and probability matching is less likely in linguistic tasks), existing studies suffer from a range of deficiencies which make it difficult to robustly assess these differences. In this paper we present three experiments which systematically test the development of probability matching behaviour over time in simple decision making tasks, across species (humans and Guinea baboons), task complexity, and task domain (linguistic vs non-linguistic). In Experiments 1 and 2 we show that adult humans and Guinea baboons exhibit similar behaviour in a non-linguistic decision-making task and, contrary to the prevailing view, a tendency to maximise (baboons) or significantly over-match (humans) rather than probability match, which strengthens over time and more so with greater task complexity; our non-human sample size ($N = 20$ baboons) is unprecedented in the probability-matching literature. Experiment 3 provides evidence against domain-specific probability learning mechanisms, showing that human subjects over-match high positive-outcome probabilities to a similar degree across linguistic and non-linguistic tasks. Our results suggest that previous studies may simply have insufficient trials to show maximising, or be too short to show maximising strategies which unfold over time. We thus provide evidence of shared probability learning mechanisms not only across linguistic and non-linguistic tasks but also across primate species.

Keywords: probability matching; comparative psychology; domain-general; decision making; language variation

Introduction

Probability matching strategies have long been thought to be characteristic of primate performance in probability learning tasks in a variety of contexts, from decision making across species to the learning of linguistic variation in humans. Probability matching occurs when subjects given probabilistic input respond in a way that is proportional to the input probabilities. However, such behaviour is not optimal in a decision theoretic sense; the optimal decision strategy is to always select the alternative with the highest positive-outcome probability, known as maximising (or regularising, in linguistic tasks). For instance, suppose one has to choose between two sources with different positive-outcome probabilities, one with positive outcomes on 70% of trials and another with positive outcomes on 30% of trials. While (full) *maximising* would secure positive outcomes 70% of the time, *probability matching* behaviour (where responses are selected in a 70-30 ratio) would lead to positive outcomes only 58% of the time ($(0.7 \times 0.7) + (0.3 \times 0.3) = 0.58$). Behaviour that does not reflect full maximising but still increases the probability of positive outcomes over probability matching is often referred to as *over-matching*—i.e., subjects over-match high positive-outcome probabilities (although not categorically).

Despite its suboptimality, probability matching behaviour has been extensively reported in decision making experiments in humans (e.g., Neimark & Shuford, 1959; Hudson Kam & Newport, 2005; Erev & Barron, 2005; however, cf. Vulkan, 2000), non-human primates (e.g., Wilson, Oscar, & Bitterman, 1964; Lau & Glimcher, 2005, 2005), and in the animal world more broadly (e.g., Bullock & Bitterman, 1962). While very few studies have directly compared behavioural differences between primate species, propensity to probability match may nevertheless differ across species: existing studies suggest that monkeys ($N = 2$ to 8) can adopt maximising strategies more readily than humans (Parrish, Brosnan, Wilson, & Beran, 2014; Brosnan, Wilson, & Beran, 2012). Moreover, there is evidence of a difference across ages in humans: while adult humans probability match in probability learning tasks, children tend to use maximising strategies instead (Derks & Paclisanu, 1967). However, this difference seems to only hold for simple binary prediction tasks: increasing the complexity of the task by introducing three or four alternatives is more likely to lead to maximising or at least increased over-matching behaviour (Gardner, 1957; Weir, 1972). Altogether, these results suggest that readily-available probability matching behaviour might be restricted to adult humans

and simple binary decision making tasks (see, e.g., Koehler & James, 2010; Vulkan, 2000).

Recent studies further suggests that the propensity to probability match might differ across domains: work comparing probability matching behaviour across linguistic and non-linguistic domains suggests that adult humans are less likely to probability match in linguistic tasks (Ferdinand, Kirby, & Smith, 2019). Nonetheless, asymmetries across ages and complexity degree remain the same in the linguistic domain: adults are more likely to probability match than children (Hudson Kam & Newport, 2005) and in tasks with fewer alternatives (Ferdinand et al., 2019). These differences across species, age groups and domains have been taken to suggest that the regularisation of linguistic variation over time might be driven by domain-specific biases as well as by domain-general and species-general biases.

In sum, despite substantial evidence on probability matching behaviour in the animal world, there are questions over the generalisability of probability matching behaviour across species, age groups, domains, and degrees of complexity (amongst other factors, see, e.g., Vulkan, 2000). However, we hitherto lack the required comparative work to robustly assess these asymmetries. In this paper we present a series of three experiments which aim to systematically test differences in behaviour in simple decision making tasks across primate species, degrees of complexity and domains. Experiments 1 and 2 compare probability matching behaviour over time in adult humans and Guinea baboons (*Papio papio*) with a hitherto unmatched sample size ($N = 20$ baboons), with tasks including different degrees of complexity (two or three alternatives). Using a similar methodology, Experiment 3 compares probability matching across domains (linguistic and non-linguistic) and degrees of complexity in humans.

Experiment 1: Maximising in Guinea baboons

Materials and Methods

We ran an experiment where, on each trial, participants were presented with geometric shapes and had to select one; different shapes had different probabilities of reward. We manipulated two factors: the number of shapes (two or three) and the reward ratios (skewed or uniform). The four conditions product of these two manipulations had the following reward ratios: 70:30 (skewed, two shapes), 70:15:15 (skewed, three shapes), 50:50 (uniform, two shapes) or 33:33:33 (uniform, three shapes). All factors were manipulated within-participant; the order of the conditions was assigned randomly per participant. All shapes were different across conditions. The preregistered design and analysis plan for this experiment is accessible at `osf.io/evxk4`.

Participants Guinea baboons (*Papio papio*) belonging to a large social group (of 25) from the CNRS Primate Center in Rousset-sur-Arc (France) participated in this study. Progress through training onto testing was conditioned on performing at criterion: participants were required to complete a condition in no more than a week. Following this criterion, we excluded the data from five baboons across conditions (final $N = 20$). Participants were 4 males (median age 4 years, min = 3, max = 12) and 16 females (median age 9 years, min = 1.5, max = 24).

The study was conducted in a facility developed by J.F. (for further information, see Fagot & Paleressompoulle, 2009; Fagot & Bonté, 2010). The baboons live in an outdoor enclosure ($700m^2$) connected to 10 computerised testing booths to which baboons have free access. Identification of the subjects within each testing booth is made possible thanks to two bio-compatible 1.2 by 0.2 cm RFID microchips implanted in each baboon's forearm. The baboons can thus participate in an experiment whenever they choose, and do not need to be captured to participate. The test program allows an independent test regime for each baboon, irrespective of the test booth it is using. Puffed rice grains are used as reward. Baboons were neither water- nor food-deprived during the research.

Procedure On each trial, participants saw a set of coloured shapes (two or three shapes, randomly positioned on a touch screen) and were prompted to select one. Each shape lead to a reward according to the ratio specified by the condition—70:30 (Skewed 2), 70:15:15 (Skewed 3), 50:50 (Uniform 2) or 33:33:33 (Uniform 3). If the participant selected the target shape for a given trial, they were rewarded (with a rice puff). If the target shape was not selected, the participant proceeded to the next trial without reward after a short delay (where a green screen signalling failure was shown). Participants completed 10 blocks of 240 trials; the reward ratios were constant across blocks within a given condition.

Analysis Our main hypotheses are that (1) subjects will maximise in the presence of a skewed reward distribution i.e. where one shape provides reward with a higher probability than the others, and (2) will do so more in the presence of three rather than two shapes as the number of probabilities to track increases and the difference between the probability of high-reward and low-reward shapes is greater.

As per the preregistered analyses, and following Ferdinand et al., 2019, we analysed the entropy drop of the set of responses (the output entropy minus the input entropy). If participants select the higher-rewarding shape more often, the variability of the set of responses will decrease thus lowering its entropy. For conditions with skewed distributions of reward we additionally analysed the choice of the optimal response (whether or not subjects choose the shape with the highest reward probability), which will confirm whether the entropy drop we observe is indeed driven by the maximisation of the number of correct predictions.

We analysed entropy drop using a linear mixed-effects model, with probability distribution of reward (skewed vs. uniform), number of shapes (three vs. two) and block as fixed effects. We included random intercepts for participant and by-participant random slopes for all fixed effects. The

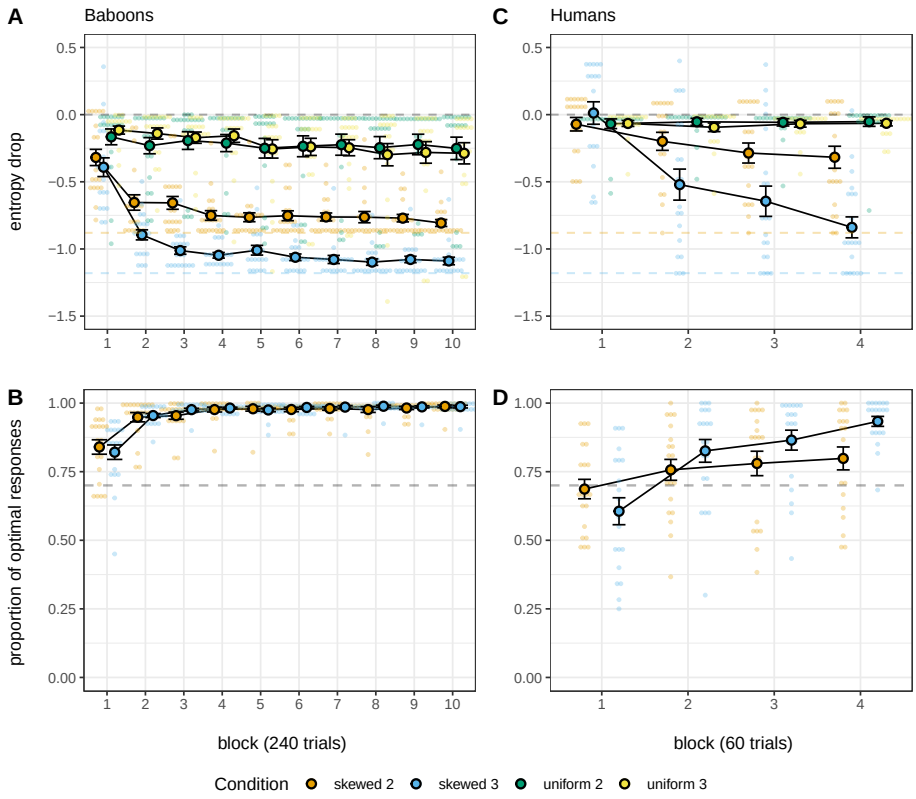

Figure 1: Summary of results for Experiments 1 (baboons, a-b) and 2 (humans, c-d). Top row: entropy drop across all four conditions in baboons (a) and humans (c). Bottom row: proportion of optimal responses in conditions with a higher-rewarding alternative (Skewed) in baboons (b) and humans (d). We show the means and standard errors as well as the individual data points (colour-faded). The grey dashed lines represent the input entropy (a–c) and the input reward probability of the higher-rewarding alternative (b–d). Coloured dashed lines in a–c represent the maximum entropy drop in Skewed 2 (orange) and Skewed 3 (blue).

proportion of optimal responses was analysed using a logistic mixed-effects model with number of shapes and block as fixed effects, by-participant random intercepts and by-participant slopes for all fixed effects. Categorical effects were simple coded so levels are compared to each other directly but the intercept is the grand mean across levels.

Results

Entropy drop A visual inspection of the results (Figure 1a) suggests that, in conditions with skewed reward distributions, the entropy of the response system decreases by block and approaches the maximum entropy drop. In contrast, entropy drop in conditions with uniform reward distributions stays around 0 throughout blocks. The results from the linear regression model confirm these observations. The model's intercept suggests a significant drop in entropy already in the first block of trials ($\beta = -0.405$, $se = 0.027$, $p < 0.001$); however, this drop is less pronounced in conditions with uniform reward distributions ($\beta = 0.244$, $se = 0.028$, $p < 0.001$)—and even less so in Uniform 3 ($\beta = 0.070$, $se = 0.011$, $p < 0.001$). We also found a significant decrease in entropy by block ($\beta = -0.028$, $se = 0.004$, $p < 0.001$), however, this decrease was not as pronounced across conditions: it is significantly flatter in uniform conditions ($\beta = 0.014$, $se = 0.002$, $p < 0.001$). The model further suggests that entropy decreases by block more with three shapes ($\beta = -0.007$, $se = 0.002$, $p = 0.001$). However, this difference is not necessarily indicative of stronger maximising behaviour with three shapes and can be explained by the differences in the maximum entropy drop across conditions (see Figure 1a).

Proportion of optimal responses Figure 1b shows the proportion of optimal responses in Skewed conditions. A visual inspection of the results suggests that the proportion of optimal responses is higher than in the input ratio of reward (70%) from the first block of responses, and that it increases with block to reach ceiling across conditions by the fourth block. The results from the logistic regression model confirm these observations. The model's intercept coefficients ($\beta = 2.231$, $se = 0.141$) suggest that participants choose the optimal responses significantly above 70% in block 1 ($z = 9.823$, $p < 0.001$), thus confirming that participants significantly over-match from very early on. The significant effect of block ($\beta = 0.451$, $se = 0.049$, $p < 0.001$) shows that the proportion of optimal responses increases further by block. We did not find a significant effect of the number of shapes ($\beta = -0.096$, $se = 0.101$, $p = 0.342$) or its interaction with block ($\beta = 0.015$, $se = 0.008$, $p = 0.083$), confirming thus

that the difference we found with entropy drop between number of alternatives was not driven by differences in maximising behaviour.

Experiment 2: Maximising in adult humans

Materials and methods

Experiment 2 adapted Experiment 1 to adult human participants. The design was as per Experiment 1 modulo the implementation of a between-participants design and the reduction of the total number of trials to 240 (to better suit the time constraints of a web experiment). The preregistered design and analysis plan for this experiment is accessible at `osf.io/b3nke`.

Participants We recruited 80 participants (N=20 per condition) through Amazon Mechanical Turk for a ten-minute long session. Participants were all over 18 years old, based in the US and had approval ratings of $> 95\%$. There were no further requirements aside from successfully completing a series of bot-screening questions to start the experiment, and finishing it in less than 50 min; no participants were excluded based on these criteria. Participants were paid a base rate of \$2 plus they received a bonus of \$0.02 (rather than a rice puff) for each correct image chosen.

Procedure The design for humans was as similar as possible to the baboons' in Experiment 1. For each trial, subjects saw two or three coloured shapes and were instructed to select one (by key press). Feedback was provided after each selection; if they selected the target image, they also received \$0.02 in bonuses. For each condition, participants went through four blocks of 60 trials each; reward ratios were constant across blocks.

Analysis We used the same models as per Experiment 1 but given the between-participant design, only intercepts for participant and by-participant slopes for the effect of block were included as random effects.

Results

Our hypotheses were the same as with baboons: (1) participants will maximise the number of correct predictions in the presence of a higher-reward alternative, and (2) they will maximise more in conditions with more alternatives (i.e., three shapes rather than two) and greater differences between high-reward and low-reward choices.

Entropy drop A visual inspection of the results (Figure 1c) suggests that, in conditions with skewed reward distributions, the entropy of responses decreases by block, although it does not reach the maximum entropy drop within 240 trials. In contrast, entropy drop in conditions with uniform reward distributions stays close to 0. The results from the linear regression model support these observations. We found a significant decrease in entropy by block ($\beta = -0.086$, $se = 0.009$, $p < 0.001$), however, this decrease was significantly smaller in Uniform conditions ($\beta = 0.089$, $se = 0.009$, $p < 0.001$). The model further suggests that entropy decreases by block more with three shapes in Skewed ($\beta = -0.047$, $se = 0.009$, $p < 0.001$) but not in Uniform conditions ($\beta = 0.046$, $se = 0.009$, $p < 0.001$); as in Experiment 1, this difference in Skewed conditions could simply be explained by the differences between maximum entropy drop.

Proportion of optimal responses Figure 1d shows the proportion of optimal responses in the Skewed conditions. A visual inspection of the results suggests that the proportion of optimal responses is not higher than in the input ratio of reward (70%) in the first block of responses, but that it increases by block. We can further observe that the increase by block is greater with three shapes. The results from the logistic regression model confirm these observations. The model's intercept coefficients ($\beta = 0.777$, $se = 0.151$) suggest that participants do not choose the optimal response significantly above 70% in block 1 ($z = -0.466$, $p = 0.642$), thus suggesting that participants' initial behaviour is not significantly different from probability matching. However, the significant effect of block ($\beta = 0.721$, $se = 0.089$, $p < 0.001$) suggests that the proportion of optimal responses increases by block to be significantly different from probability matching from the second block onward ($p < 0.001$). Results further show that the increase in optimal responses is greater in Skewed 3 ($\beta = 0.371$, $se = 0.088$, $p < 0.001$) thus confirming a stronger maximising behaviour with three rather than two shapes.

Cross-species comparison So far results suggest that both baboons and humans ultimately use maximising and over-matching rather than probability matching strategies to solve binary and ternary prediction tasks. However, results across species cannot be straightforwardly compared given the differences in the number of trials per testing block (240 for baboons and 60 for humans). This difference was due to an overestimation, based on previous research (e.g., Wilson et al., 1964), of the amount of trials it would take baboons to reach a maximising strategy. We can still nevertheless calculate the difference between the empirical probabilities of reward in sub-blocks of 60 trials and the proportion of optimal responses within these sub-blocks. For humans, the empirical reward probabilities are always 0.7 because by design each block of 60 trials rewarded the optimal response at exactly that rate. For baboons, these probabilities will vary slightly across sub-blocks of 60 trials, given that the reward probabilities were controlled with respect to longer 240-trial blocks. To resolve this, we calculate the difference between participants' empirically-observed reward rates and their optimal response rates for each 60-trial block: a difference of 0 between input and output proportions of optimal responses would mean that the choice of optimal responses corresponds perfectly to the input probabilities of reward (i.e. probability matching); differences above 0 indicate that the optimal response is chosen more often than its proportion of reward (i.e. over-matching or maximising behaviour).

Figure 2 shows the increase in the proportion of optimal

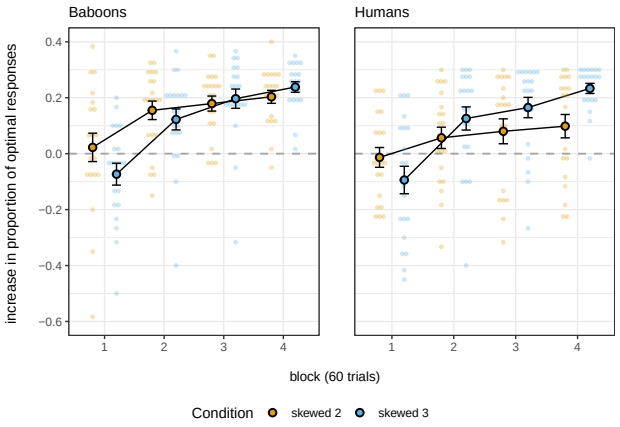

Figure 2: Increase in the proportion of optimal responses in Skewed conditions (Experiments 1–2) in baboons and humans within the initial 240 trials, broken down into blocks of 60 trials. Means and standard errors as well as individual data points are shown. Dashed lines represent the input reward probability (i.e., a difference of 0 from it).

responses within the first 240 trials across both species, segmented in blocks of 60 trials. We ran a linear regression model with fixed effects for species (baboons vs humans), the number of shapes (three vs two), and block (of 60 trials each). As random effects we included intercepts for participant and by-participant random slopes for the effects of block and the number of shapes. Results suggest that the difference between input and output proportions of optimal responses was not different from 0 initially ($\beta = -0.005$, $se = 0.020$, $p = 0.813$) but that it increased significantly by block ($\beta = 0.074$, $se = 0.007$, $p < 0.001$); the non-significant effect of species along with the non-significant interaction between species and block suggest that this increase is comparable across species ($\beta = 0.017$, $se = 0.020$, $p = 0.403$; $\beta = 0.005$, $se = 0.007$, $p = 0.474$). We further found a significant interaction between block and the number of variants suggesting that the proportion of optimal responses increases more by block in conditions with three shapes ($\beta = 0.028$, $se = 0.006$, $p < 0.001$); moreover, we did not find a three way interaction with species ($\beta = -0.006$, $se = 0.006$, $p = 0.347$), suggesting that this difference holds across species as well. In other words, baboons and humans performed in the same way on this task, both exhibiting an early phase of exploration that resembles probability-matching followed by movement towards maximisation.

Experiment 3: Over-matching across domains

Experiments 1 and 2 provide evidence against probability matching behaviour in humans and baboons: when a maximising strategy is available, both species show a tendency towards maximisation over time, across different degrees of complexities. In Experiment 3 we test whether humans are even more likely to over-match in linguistic tasks compared to non-linguistic tasks (as suggested in Ferdinand et al., 2019). We use a similar design to that in Experiment 2 but adapt it to be comparable to Ferdinand et al. (2019), which has previously shown differences between domains. Participants in Ferdinand et al. (2019)'s non-linguistic tasks are asked to choose between differently coloured marbles with different probabilities of occurrence; in their linguistic tasks, participants are asked to choose between linguistic variants (alternative forms to convey the same meaning). In common with much of the literature on linguistic probability matching/regularisation, these tasks are divided into training and testing phases. During training participants are *passively exposed* to a set of shapes/words with different probabilities of occurrence; in testing, they are asked to predict a sequence of the same shapes/words *without feedback*. In Experiment 3 we adapt our Experiment 1–2 method to more closely resemble these tasks: we eliminate feedback and monetary reward during training; during testing, participants are told that their responses will influence their monetary reward, but do not receive trial-by-trial feedback which would allow them to adjust their behaviour accordingly.

Materials and Methods

We ran a between-participants experiment where we manipulated two variables: domain (words or shapes) and number of variants (two or three). The reward distribution was skewed across all conditions: the input reward ratios were 70:30 and 70:15:15 for two and three variants respectively as in Experiments 1–2. In the non-linguistic (shapes) conditions participants were exposed to a sequence of shapes and then were asked to predict a sequence of the same shapes themselves. In the linguistic (words) conditions participants were exposed to a sequence of words (two or three non-words which were used variably to refer to a single referent object, i.e., the words were synonyms which were deployed unpredictably, as in, e.g., Ferdinand et al., 2019) and were then asked to predict a sequence of words produced by the same population of speakers. Crucially, participants were told they would be rewarded for each correct prediction in the sequence they produced, but unlike in Experiments 1–2 they did not receive feedback on each trial.

Participants As per Experiment 2, we recruited 104 English-speaking adult participants for a ten-minute long session. Participants were evenly distributed amongst the four conditions ($N = 25$ across all conditions except for Shapes-Skewed 2, where $N = 29$).

Procedure The experiment was divided into a training and a testing phase. During training participants were exposed to a sequence of shapes or words. One of the shapes/words appeared more often than the other: the input ratios of occurrence were 70:30 and 70:15:15 for conditions with two (Skewed 2) and three variants (Skewed 3) respectively. During testing, participants were asked to predict a sequence of shapes/words based on their training. Participants went through two blocks of 60 trials during training and two blocks of 60 trials during testing; this makes the total number of tri-

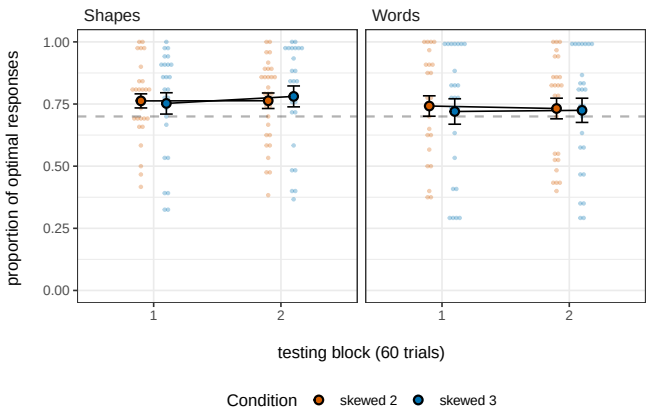

Figure 3: Proportion of optimal responses (i.e., majority variants in training) during testing trials in Experiment 3, across domains (words or shapes) and degrees of complexity (two or three shapes/words). Means and standard errors as well as individual data points are shown. Dashed lines represent the input probability of occurrence during training.

als equivalent to Experiment 2.

Analysis We analysed a single outcome variable: the choice of optimal responses. We ran the same model as per Experiment 2 but with an additional fixed effect for domain (words vs shapes).

Results

Figure 3 shows the proportion of optimal responses. A visual inspection of the results suggests that the proportion of optimal responses (i.e., majority variant) in testing is only slightly higher than in training (70%) on average. We do not observe any obvious differences between blocks or between domains. The results from the logistic regression model show that participants choose optimal responses subtly but significantly above 70% in testing block 1 ($\beta = 1.557$, $se = 0.160$; $z = 4.432$, $p < 0.001$), thus suggesting that participants' subtle over-matching behaviour is significantly different from probability matching. We found no other significant effects, suggesting similar behaviour across blocks ($\beta = 0.069$, $se = 0.067$, $p = 0.304$), number of variants ($\beta = 0.027$, $se = 0.158$, $p = 0.866$), and domains ($\beta = 0.063$, $se = 0.159$, $p = 0.693$).

Discussion

Our results provide evidence against probability matching behaviour in simple decision making tasks in baboons as well as in adult humans. In Experiment 1 and 2, where the distribution of reward was skewed (i.e. 70:30 or 70:15:15) and a maximising strategy was therefore available, both species showed an initial exploratory behaviour resembling probability matching followed by a switch to over-matching behaviour within 240 trials: in the first block of 60 trials, the selection of the shape with the highest reward probability was not significantly different from its reward probability, but there was a significant increase of over-matching by block. In Experiment 1, where we were able to collect 2400 trials per baboons, we observe a convergence on maximising behaviour after 240 trials. We did not observe the same convergence in Experiment 2 with humans because we did not collect as many trials for humans in Experiment 2, but given that we found comparable behaviour within 240 trials across species, we predict that over-matching behaviour in humans would increase by block after the 240 trials in a similar fashion to eventually converge on maximising; we are currently running a longer version of the same experiment to test this. Crucially, we also found probability matching behaviour in both species when the reward distribution was uniform. This difference in behaviour between skewed and uniform conditions suggests that maximising is not the default strategy but that both species are sensitive to the availability of maximising strategies.

We also found evidence for an effect of degree of complexity: an earlier tendency towards maximising behaviour in ternary rather than binary prediction tasks, where not only the number of choices increased but also the differences between probabilities of reward for higher-reward and lower-reward variants was greater. In the results obtained from baboons, we observe that this difference is later lost as participants converge towards maximising behaviour after the initial 240 trials. Altogether, regardless of the tendency towards maximisation, it is still possible that the initial period of exploration is longer than expected from a decision theoretic sense. Further work is required to assess differences in the trade-off between exploration and exploitation with a more fine grained analysis of decision making over time across species and different reward distributions (for differences across age groups, see Sumner et al., 2019).

Finally, we did not find evidence of a difference in human behaviour across our non-linguistic and linguistic tasks (cf. Ferdinand et al., 2019). We found that participants slightly over-produce the majority input variant, thus suggesting a weak tendency toward over-matching across domains. However, we did not observe as strong a tendency towards maximising behaviour in Experiment 3 as in Experiment 2, which could be due to the lack of feedback and the passive exposure to the input probabilities. It is possible that the lack of reinforcement during exposure hinders the learning of the input probabilities. Further, the lack of feedback to participants' responses at each trial during testing could also impact their selection strategies: without feedback, participants might be more likely to focus on the prediction of an entire sequence of outcomes (and not of a single outcome at a time) which in turn may increase the probability of the production of rare variants. It is nevertheless worth noting that over-matching behaviour in Experiment 2 is not as strong in block two as in the following blocks, suggesting weak over-matching behaviour within 120 trials even with feedback and reward. This suggests that the presence of feedback and reward is required to produce the over-matching and maximising behaviour we see in Experiments 1–2; it does not sim-

ply emerge from repeated testing. Interestingly, there is some variability in the experimental methods used to demonstrate probability matching / regularisation in the experimental literature in linguistics: some studies avoid feedback in testing (e.g. Ferdinand et al., 2019) whereas others provide feedback on at least some test trials (e.g. Culbertson, Smolensky, & Legendre, 2012), which may induce (subtle) differences in the tendency to maximise. It is also worth noting that the over-matching effect we observe in Experiment 3 is small: participants only over-produce the majority variant by about 5%, that is, producing it on 75% of trials rather than 70% as in their input. Since most experiments on linguistic probability matching use far fewer test trials (e.g. 10 trials in Ferdinand et al., 2019), they will typically lack the statistical power and/or response granularity to differentiate between this small level of over-matching and probability-matching.

Conclusion

Our results provide evidence against the common assumption that monkeys and humans are likely to probability match in simple decision making tasks and raise questions over the validity of conclusions in standard behavioural experiments, which our results suggest may simply have insufficient trials to show a maximising behaviour which unfolds over time. It also casts doubt on the suggested domain-specific sources of maximising behaviour in linguistic tasks by providing evidence of shared mechanisms in probability learning not only across primate species but across linguistic and non-linguistic tasks.

Acknowledgements This work was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 681942, held by KS).

Ethics The study with baboons received approval from the Animal Welfare and Ethical Review Body of the University of Edinburgh (Ref # OS5-19), and the French Ministère de l'Education Nationale et de la Recherche (approval # APAFIS-2717-2015111708173794-V3). The studies with humans were approved by the PPLS Research Ethics Committee at the University of Edinburgh (Ref # 378-1819/1).

References

Brosnan, S. F., Wilson, B. J., & Beran, M. J. (2012). Old world monkeys are more similar to humans than new world monkeys when playing a coordination game. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1733), 1522–1530. doi: 10.1098/rspb.2011.1781

Bullock, D. H., & Bitterman, M. E. (1962). Probability-Matching in the Pigeon. *Am. J. Psychol.*, *75*(4), 634–639. doi: 10.2307/1420288

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329. doi: 10.1016/j.cognition.2011.10.017

Derks, P. L., & Paclisanu, M. I. (1967). Simple strategies in binary prediction by children and adults. *Journal of Experimental Psychology*, *73*(2), 278. doi: 10.1037/h0024137

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychol. Rev.*, *112*(4), 912–931. doi: 10.1037/0033-295X.112.4.912

Fagot, J., & Bonté, E. (2010). Automated testing of cognitive performance in monkeys: Use of a battery of computerized test systems by a troop of semi-free-ranging baboons (papio papio). *Behavior research methods*, *42*(2), 507–516. doi: 10.3758/BRM.42.2.507

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behavior Research Methods*, *41*(2), 396–404. doi: 10.3758/BRM.41.2.396

Ferdinand, V., Kirby, S., & Smith, K. (2019). The cognitive roots of regularization in language. *Cognition*, *184*, 53–68. doi: 10.1016/j.cognition.2018.12.002

Gardner, R. A. (1957). Probability-learning with two and three choices. *The American Journal of Psychology*, *70*(2), 174–185. doi: 10.2307/1419319

Hudson Kam, C. L., & Newport, E. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Lang. Learn. Dev.*, *1*(2), 151–195. doi: 10.1207/s15473341lld0102_3

Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Mem. Cognit.*, *38*(6), 667–676. doi: 10.3758/MC.38.6.667

Lau, B., & Glimcher, P. W. (2005). Dynamic Response-by-Response Models of Matching Behavior in Rhesus Monkeys. *J. Exp. Anal. Behav.*, *84*(3), 555–579. doi: 10.1901/jeab.2005.110-04

Neimark, E. D., & Shuford, E. H. (1959). Comparison of predictions and estimates in a probability learning situation. *Journal of Experimental Psychology*, *57*(5), 294. doi: 10.1037/h0043064

Parrish, A. E., Brosnan, S. F., Wilson, B. J., & Beran, M. J. (2014). Differential Responding by Rhesus Monkeys (Macaca mulatta) and Humans (Homo sapiens) to Variable Outcomes in the Assurance Game. *Anim. Behav. Cogn.*, *1*(3), 215–229. doi: 10.12966/abc.08.01.2014

Sumner, E., Li, A. X., Perfors, A., Hayes, B., Navarro, D., & Sarnecka, B. W. (2019). The exploration advantage: Children's instinct to explore allows them to find information that adults miss. *PsyArXiv*. doi: 10.31234/osf.io/h437v

Vulkan, N. (2000). An economist's perspective on probability matching. *J. Econ. Surv.*, *14*(1), 101–118. doi: 10.1111/1467-6419.00106

Weir, M. W. (1972). Probability performance: Reinforcement procedure and number of alternatives. *The American Journal of Psychology*, 261–270. doi: 10.2307/1420666

Wilson, W. A., Oscar, M., & Bitterman, M. E. (1964). Probability-learning in the monkey. *Q. J. Exp. Psychol.*, *16*(2), 163–165. doi: 10.1080/17470216408416361