TITLE

Persistent Decision-Making in Mice, Monkeys, and Humans

AUTHORS:

Veldon-James Laurie[1*], Akram Shourkeshti[1], Cathy S. Chen[2], Alexander Herman[3], Nicola Grissom[2], R. Becket Ebitz[1]

1. Department of Neuroscience, University of Montreal, Quebec, Canada
2. Department of Psychology, University of Minnesota, Minneapolis, United States
3. Department of Psychiatry, University of Minnesota, Minneapolis, United States

Corresponding Author:
Veldon-James Laurie*
*Corresponding Author: veldonjames.laurie@gmail.com

ABSTRACT

Humans have the capacity to persist in behavioural policies, even in challenging environments that lack immediate reward. Persistence is the scaffold on which many higher executive functions are built. However, it remains unclear whether humans are uniquely persistent or, instead, if this capacity is widely conserved across species. To address this question, we compared humans with mice and monkeys in harmonised versions of a dynamic decision-making task. The task encouraged all species to strike a balance between persistently exploiting one policy and exploring alternative policies that could become better at any moment. Although all three species had similar strategies, we found that both primate species—humans and monkeys—were able to persist in exploitation for longer than the mice. The similarities in persistence patterns in humans and monkeys, as opposed to mice, may be related to the various ecological, neurobiological, or cognitive factors that differ systematically between these species.

TEASER

Humans, monkeys and mice use similar decision-making strategies but differ in their persistence in exploitation.

INTRODUCTION

Decision-making in an uncertain environment requires a fine balance between two goals. Decision-makers must persist in exploiting previously rewarded options, but also regularly explore alternatives that have the potential to be even better. In humans, exploratory decision-making drives our everyday interactions (Rich and Gureckis, 2018), problem resolutions (Knox et al., 2012), goal achievements (Wilson et al., 2021) and predicts individual differences in self-reported engagement (Yan et al., 2023). However, our capacity to strike a balance between exploitation and exploration is also fragile. The balance is easily thrown off by stress (Kaske et al., 2023) and drug addiction (Verdejo-García et al., 2006) and is dysregulated in many neurological conditions, like obsessive-compulsive disorder (Tolin et al., 2009), depression (Blanco et al., 2013), anxiety (Teng et al., 2016), and ADHD (Mäntylä et al., 2012). Because evolution tends to canalise phenotypes over time (Waddington, 1942; Siegal and Bergman, 2002)—making behaviour more robust against influence of environmental and developmental perturbations—these results could suggest that the human capacity to balance exploration and exploitation may have evolved relatively recently. However, in part because of the difficulty of harmonising tasks and data collection across species, we do not know how exploratory decision-making in humans compares against other species.

There is urgent need for comparative analyses of human and non-human exploratory decision-making as animal models are increasingly being used to model human decision-making. This is most obvious in the mouse, where the rise of optogenetics and other techniques dependent on genetic expression (Boyden et al., 2005) means that mice have been increasingly involved in cognitive function research in recent years (Ellenbroek and Youn, 2016). This is especially true in the area of decision-making under uncertainty, where there has been an explosion of research using rodent models (Saddoris et al., 2015; Groman et al., 2016; Bari et al., 2019; Izquierdo et al., 2019; Soltani and Izquierdo, 2019; Chen et al., 2021a, 2021b; Grossman et al., 2022; Iyer et al., 2022). Critically, this work is already beginning to find important phenomenological and computational differences between species in these tasks (Trepka et al., 2021; Woo et al., 2023), though it remains unclear where humans fit within these frameworks. Ultimately, translating findings from animal studies into understanding and applications in humans will require comparative studies (Manger et al., 2008; Stevenson et al., 2018; Woo et al., 2023), which can uncover the variability, similarities, and differences within and across species by contrasting their strategies in tasks.

Here, we asked if human patterns of exploratory decision-making are unique or else shared with other species. We focused on comparing humans against two of the most commonly used animal models in psychology and cognitive neuroscience: Mus musculus (the mouse) and Macaca mulatta (the rhesus monkey). Because these species deviated from the human lineage at different times (monkey: 23-25 million years ago (Disotell and Tosi, 2007; Gibbs et al., 2007); mice: ~90 million years ago (Ernst and Carvunis, 2018)), we reasoned that any feature of exploratory decision-making that was unique to humans would most likely have evolved within the last 23-25 million years (or else been lost over time in one or both of the other species). Conversely, any feature that was shared between all three species would most likely have evolved over roughly

90 million years ago (or else proven so adaptive that it independently evolved in all three species via convergent evolution).

To identify the similarities and differences between humans, mice, and monkeys, all three species performed harmonised versions of a classic explore/exploit task known as a restless multi-armed bandit. In this task, participants are presented with a series of trials where they had to make choices between identical targets. Each target's reward probability changes independently and unpredictably over time. In consequence, all three species confront the same conundrum: should they persist in exploiting an already rewarding option or should they explore alternative options? The tasks were "harmonised" in the sense that every species performed a spatial bandit task in which the key parameters of the reward schedules were matched across species—making the structure of the task as similar as possible for mice, monkeys and humans—even as they reported their choices using species-typical experimental apparatuses. Although all three species exhibited the behavioural signatures of exploration and exploitation, there were large differences in how the species switched between targets. Computational modelling revealed that the key difference between mice (who switched frequently) and monkeys and humans (who did not) lay in the primate species' capacity to persistently exploit options for much longer than the mouse. Control experiments in humans ruled out several low-level explanations for these species' differences. Together, these results suggest that the primate lineage may have only recently evolved its remarkable capacity to persist in exploitative states.

RESULTS

In **Experiment 1**, mice (N=32, average of 7.97 sessions each, total of S = 255 sessions, 70,478 total trials), monkeys (N=5, average of 18.6 sessions each, total of S = 93 sessions, 57,878 trials) and humans (N=258, 1 session each, total of S = 258 sessions, 77,400 total trials) performed comparable spatial restless k-armed bandit tasks (**Figure 1A**). Each target offered a probability of reward which changed slowly and independently over time (**Figure 1B**). As a result, the task encouraged participants to both exploit rewarding targets and explore new targets to learn about other potential rewards. Mice indicated their choices via nose pokes, monkeys via saccadic eye movements, and humans with a computer mouse (**Figure 1A**). We controlled for species differences in the timing of the task, the number of targets and the response modality via 3 additional experiments in humans (**see Methods**; **Experiment 2**, **Experiment 3, and Experiment 4**).

There were differences in performance between species: humans were more likely to get rewards compared to mice, but not monkeys (normalised difference from chance; **Figure 1C**; humans-mice: $p > 0.0001$, $b = -13.04$, 95% CI = [-17.78, -8.29], $t(511) = -5.40$; humans-monkeys: $p = 0.24$, $b = -6.66$, 95% CI = [-17.77, 4.46], $t(349) = -1.18$). Monkeys also performed better than mice ($p < 0.0001$, $b = –7.65$, 95% CI = [–10.94, –4.37], $t(344) = -4.59$). Humans switched less often than mice (**Figure 1D**; $p > 0.0001$, $b = 0.19$, 95% CI = [0.12, 0.20], $t(511) = 7.21$), as did monkeys ($p < 0.0001$, $b = 0.26$, 95% CI = [0.18, 0.34], $t(344) = 6.17$). There were no differences in switch

probability between humans and monkeys (**Figure 1D**; p = 0.33, b = -0.056, 95% CI = [CI [-0.17, 0.057], t(349) = -0.98). In sum, there were species differences in both task strategy and outcome, driven by differences between primate species and the mice.

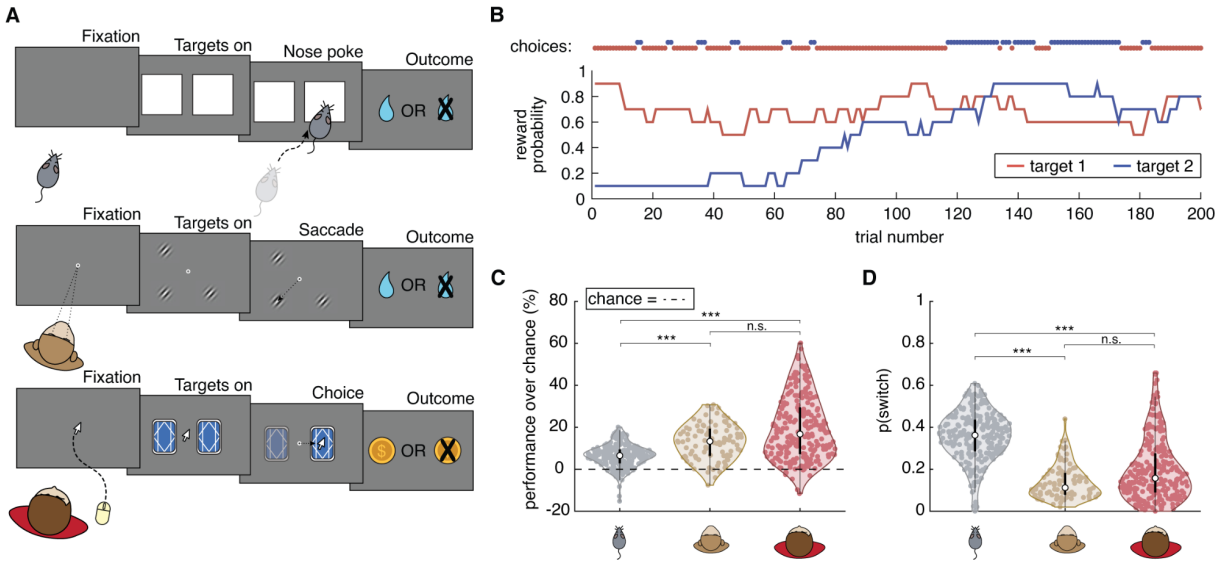


Figure 1. Task design and behaviour across species. **A**) A schematic representation of the bandit task in each species (mice = top, monkeys = middle, humans = bottom). **B**) Example reward schedule, including 200 trials from one session with one human. The reward probabilities of each of the 2 targets (blue and red traces) walk randomly, independently across trials. The humans' choices are illustrated as coloured dots along the top. **C**) Percentage of reward relative to chance in all species. Thick black lines = IQR, thin = whiskers, open circle = median. Black dotted line = chance performance. **D**) Probability of switching targets during the task between species. Same conventions as C. Asterisks represent significance levels as follows: * indicates standard α = 0.05 threshold; ** indicates α = 0.001; and *** indicates α = 0.0001, corrected for multiple comparisons.

Switching dynamics and exploratory behaviour

Switching happens for multiple reasons in this task (Ebitz et al., 2018; Chen et al., 2021b). Sometimes animals switch options because they are engaging in rapid trial and error sampling. Other times they switch because the option they have been choosing is no longer rewarding. To determine how these distinct types of switching differed across all species, we fit a "mixture model" to the distribution of inter-switch intervals (number of trials between switches) in each species (**Figure 2A**; **see Methods** for more details).

We found that the behaviour of all species could be best described as a mixture of two switching modes (**Figure 2A, Supplementary Table 1**). Participants sometimes switched between targets at a fast pace ("switching regime") and they sometimes stuck to choosing one target repeatedly ("persistent regime"). Humans and monkeys differed from mice in their average switching probability during the persistent regime (**Figure 2B**; humans-mice: p < 0.0001, b = 0.081, 95% CI = [0.056, 0.11], t(500) = 6.32; monkeys-mice: p < 0.0001, b = 0.16, 95% CI = [0.11, 0.22], t(341) = 5.77), though there were no differences between primates (humans-monkeys : p = 0.29, b = -0.056, 95% CI = [-0.16, 0.047], t(341) = -1.063). Humans also differed from both monkeys and mice in the average switching probability during the switching regime (humans-mice: p < 0.0001,

b = 0.19, 95% CI = [0.14, 0.24], t(500) = 6.94; humans-monkeys: $p < 0.017$, b = 0.15, 95% CI = [0.034, 0.27], t(341) = 2.53), whereas monkeys did not differ from mice here ($p = 0.28$, b = 0.054, 95% CI = [-0.044, 0.15], t(341) = 1.083). No species differed from each other in the relative weight of the two regimes (**Supplementary Table 2**). Together, these results suggest that species differences in switching in **Figure 1B** were largely driven by the primates' increased tendency to persist, compared to mice.

In order to determine why primates switched less frequently during the persistent regime, we categorised individual choices based on the underlying reason for those decisions. Specifically, we used a Hidden Markov Model (HMM; **Figure 2C**; **see Methods**) to infer whether individual choices were more likely to be due to a state of exploratory, trial- and-error sampling or a state of exploitative choices to a single option (Ebitz et al., 2018, 2019; Chen et al., 2021b; Kaske et al., 2023). Example choice sequences, with labels, are in **Figure 2D**. Based on the HMM labels, the probability of exploring differed between both humans and monkeys, compared to mice (**Figure 2E**; humans-mice: $p < 0.0001$, b = 0.29, 95% CI = [0.22, 0.35], t(479) = 8.82; monkeys-mice: $p < 0.0001$, b = 0.50, 95% CI = [0.33, 0.67], t(327) = 5.82). Again, there were no differences within primates (humans-monkeys: $p = 0.22$, b = -0.13, 95% CI = [-0.33, 0.076], t(334) = -1.24). These results suggest that primates switched less on average because they were less exploratory than mice.

One reason for the decrease in exploration in primates compared to mice could be a change in the stability of explore and exploit states across species. To determine if there were differences in the stability of these states, we analysed the parameters of the fitted HMM (Ebitz et al., 2018, 2019; Chen et al., 2021b). We found species differences in the likelihood of staying in exploitation (exploit-to-exploit transition probability: humans-mice: $p < 0.0001$, b = -0.092, 95% CI = [-0.12, -0.067], t(479) = -6.98; humans-monkeys: $p = 0.32$, b = 0.055, 95% CI = [-0.054, 0.16], t(334) = 1.00; monkeys-mice: $p < 0.0001$, b = -0.16, 95% CI = [-0.22, -0.095], t(327) = -5.00). Specifically, mice were less likely to stay in exploitation than either primate (mean mice: 0.77 ± 0.19 STD across sessions, monkeys: 0.92 ± 0.046 STD, humans: 0.86 ± 0.13 STD). This suggests that primates persisted more in exploitation compared to mice.

The fitted HMMs represent a system of equations describing the dynamics of behaviour in each session, therefore we next performed mathematical analyses of these equations in order to determine how the dynamics of exploration and exploitation changed across species (**Figure 2F**; see **Methods**). Specifically, we considered (1) the difference in potential energy between exploration and exploitation (a measure of the difference in the stability of the two states), and (2) the activation energy needed to transition from exploitation to exploration (a measure of the stability of exploitation). In mice, we found that exploration and exploitation had roughly the same level of potential energy (mean difference in energy = 0.19 ± 1.55 STD across sessions). By contrast, exploitation tended to be a deeper, more energetically stable state than exploration in both primates (monkeys: -1.87 ± 0.72 STD; humans: -1.30 ± 1.61 STD). Both primates were significantly different than mice (humans-mice: $p < 0.0001$, b = 1.47, 95% CI = [1.053, 1.88], t(479) = 6.95; monkeys-mice: $p < 0.0001$, b = 2.29, 95% CI = [1.30, 3.27], t(327) = 4.56), but not different from each other (humans-monkeys: $p = 0.38$, b = -0.58, 95% CI = [CI [-1.87, 0.72], t(334) = -0.88).

The amount of energy required to end exploitation was lower in mice than humans, whereas humans and monkeys did not differ (differences in activation energy in humans-mice: $p < 0.0001$, $b = -0.74$, 95% CI = [-1.08, -0.40], $t(479) = -4.30$; humans-monkeys: $p = 0.84$, $b = 0.13$, 95% CI = [-1.13, 1.39], $t(334) = 0.20$; mean across species in mice = 1.77 ± 1.39 STD, monkeys = 2.69 ± 0.57 STD, humans = 2.54 ± 1.53 STD). Monkeys showed only a trend level difference in this measure compared to mice that did not survive correction for multiple comparisons (monkeys-mice: $p = 0.029$, $b = -0.083$, 95% CI = [-1.58, -0.084], $t(327) = -2.19$). There were also no differences in performance or the probability of switching or exploration with either self-reported gender in humans or sex in mice (**Figure S1**; all monkeys were male). In short, primates—both humans and monkeys—exhibited deeper, more stable exploitative states compared with mice, consistent with their lower switching (**Figure 1D**) and exploration rates (**Figure 2B**).

<u>Impulsivity and satiety do not explain mouse switching dynamics</u>

It is possible that mice were less persistent in exploitation because they were generally more "impulsive" than other species. In this view, the greater frequency of "exploratory" choices in the mice would be due to impulsive decisions in this species. We reasoned that "impulsive" choices would be faster than other kinds of choices—in contrast to genuine exploratory choices that tend to be slower (Ebitz et al., 2018; Chen et al., 2021b). Therefore, to determine if there were species differences in impulsivity, we analysed the reaction times in the explore/exploit trials (**Figure 2G, top**). To permit comparison across species with differing response modalities (i.e. touch screen for mice, saccades for monkeys, and computer mouse for humans), we focused on how response time was modulated across states (see **Methods**). We found that all 3 species, mice included, had a mean modulated index which was below zero for exploration / exploitation, which meant that all three species had slower reaction times for exploration compared to exploitation (**Figure 2H - top;** different from zero in mice: $p < 0.0001$, $b = -0.15$, 95% CI = [-0.19, -0.099], $t(222) = -6.08$, monkeys: $p < 0.0002$, $b = -0.51$, 95% CI = [-0.75, -0.26], $t(78) = -4.093$, humans: $p < 0.0001$, $b = -0.28$, 95% CI = [-0.34, -0.23], $t(238) = -10.52$). This implies that exploration was a deliberative, rather than impulsive choice, in all three species, including the mice.

Although exploratory choices were not more impulsive in mice as a rule, it remained possible that the lower levels of performance in mice were due to impulsivity. Therefore, we next performed the same response time analysis for rewarded versus unrewarded trials (**Figure 2G, bottom**). We found that all species had, on average, slower reaction times on unrewarded trials compared to rewarded trials (**Figure 2H - bottom;** different from zero in mice: $p < 0.0001$, $b = -0.071$, 95% CI = [-0.088, -0.054], $t(254) = -8.21$; monkeys: $p < 0.001$, $b = -0.079$, 95% CI = [-0.13, -0.034], $t(78) = -3.48$; humans: $p < 0.0005$, $b = -0.040$, 95% CI = [-0.062, -0.018], $t(257) = -3.57$), consistent with the idea that all 3 species had speeded decisions when they were about to be correct, not incorrect. This suggests that unrewarded trials were not caused by impulsive, overly fast choice processes in any of the species, including mice. Together, these results suggest that mice had a strategy that was no more impulsive than either primate species even when that strategy was less rewarding and produced more exploratory choices.

Despite the species' similar patterns of response time across exploration and exploitation and rewarded and unrewarded trials, it remained possible that mice explored more because they became more satiated over the course of the session than monkeys and humans. In this view, the similar patterns of response time modulation in mice were just an artifact of averaging across the entire session. Therefore, we examined how several reaction time and behavioural metrics changed within each session within each species. Examining the reaction times of the first 150 and last 150 trials for each species revealed no species-specific increase in reaction times over time. All three species sped up over the first ~20 trials but then were largely stable over the last 150 trials (**Figure S2A**). Critically, there was no increase in the probability of switching in the persistent regime in any species (**Figure S2B,** paired t-test in mice: $p = 0.205$, 95% CI = [-0.44, 2.05], $t(245) = 1.27$; monkeys: $p = 0.151$, 95% CI = [-1.24, 7.90], $t(86) = 1.45$; humans: $p = 0.871$, 95% CI = [-3.36, 2.85], $t(214) = -0.16$). Further, mice, specifically, did not explore more towards the end of the session (**Figure S2C;** $p=0.843$, 95% CI = [-0.021, 0.026], $t(237)=0.20$), though they also did not demonstrate the slight increase in exploitation seen in both primates (**Figure S2C**, see **Methods**; monkeys: $p < 0.005$, 95% CI = [0.014, 0.072], $t(92) = 2.92$; humans: $p < 0.001$, 95% CI = [0.024, 0.074], $t(242) = 3.82$). Considering that monkeys were working for primary reinforcement and humans for monetary reinforcement, the increase in exploitation in both species is unlikely to be a simple consequence of satiety. In sum, satiety did not appear to be a sufficient explanation for the increase in exploration in mice, nor a major driver of switching behaviour in any species.

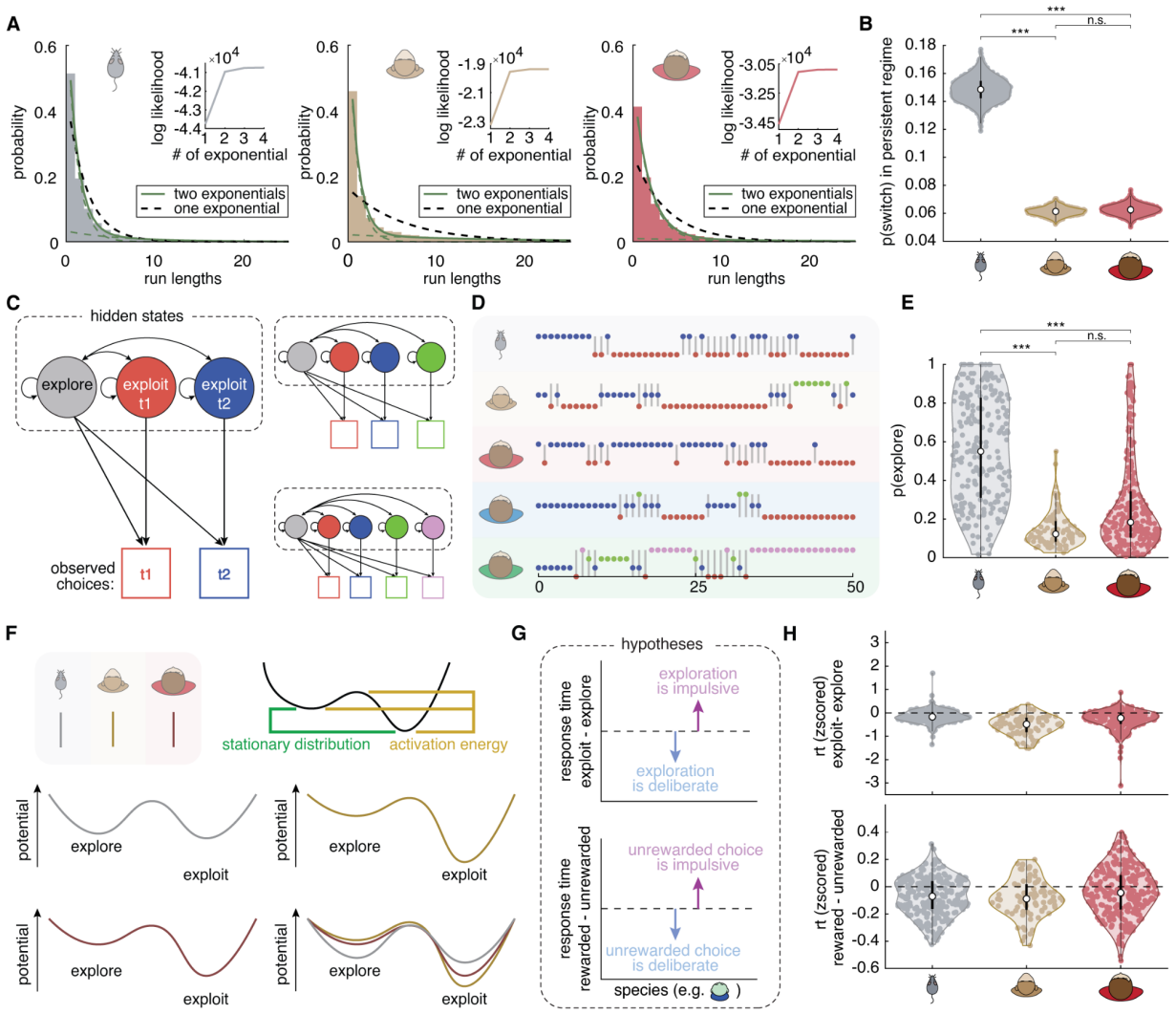


Figure 2. Different patterns of switching and exploration across species. A) Distributions of the number of trials between switch decisions ("run lengths") in mice, monkeys and humans. If the species had a fixed probability of switching, run lengths would be exponentially distributed (black dotted line). A mixture of two exponential distributions (purple line) suggests 2 distinct probabilities of switching. Dotted purple lines show each mixing distribution, one slow-switching and another fast-switching. (Inset) Log likelihoods for different mixture models containing a range of 1 to 4 exponential distributions in each species. **B**) Bootstrapped estimates of switch probability for the slow-switching distribution (the "persistent regime") across species. Thick black lines = IQR, thin = whiskers, open circle = median. **C**) Hidden Markov models (HMMs) were used to infer the goal state on each trial from the sequence of choices. The model included one persistent state for each target (''exploit'') and one state in which subjects could choose any of the targets ("explore"). Right) The model can be extended to account for different numbers of targets by adding exploit states. **D**) Fifty-trial example choice sequences for mice, monkeys and humans. The coloured circles represent the chosen target, and the grey lines highlight the explore choices identified with the HMM. **E**) Probability of exploration across species, as inferred by the HMM. Same conventions as B. **F**) Fitting the HMM involves identifying a set of equations that describe the dynamics of exploration and exploitation, meaning the rate at which participants explore, exploit, and switch between states. Left) Certain analytic measures of these equations, namely their stationary distributions (Boltzmann, 1868) and activation energies (Arrhenius, 1889) can be used to derive an intuitive picture of the landscape of state dynamics. Middle) Average state dynamic landscapes for each species. Right) State dynamic landscapes for all species overlaid G) Differences in response time (RT) can reveal which choices are fast and impulsive or slow deliberative. Top: a cartoon illustrating RT differences under the hypothesis that exploration is more deliberative than exploitation

(blue) or that exploration is more impulsive (purple). Bottom: same for rewarded and unrewarded trials. H) Z-scored RT differences across all three species. Top: difference across HMM-defined exploration and exploitation. Bottom: between rewarded and unrewarded trials. Thick black lines = IQR, white circle = median. Asterisks represent significance levels as follows: * indicates standard α = 0.05 threshold; ** indicates α = 0.001; and *** indicates α = 0.0001, corrected for multiple comparisons.

<u>Manipulating task variables to understand species differences</u>

Although there were species differences in performance and persistence, these could be artifacts of the small variations in task design across species. For example, 2 of the monkeys experienced reward walks that were slightly richer on average than the other monkeys, mice, and humans. Still, this did not appear to impact the probability of switching (monkey Normal Env. vs monkey Rich Env.: p=0.77, b = -0.013, 95% CI = [-0.099, 0.073], t(89) = -0.29), nor the probability of exploration (monkey Normal Env. vs monkey Rich Env.: p = 0.91, b = 0.0051, 95% CI = [-0.087, 0.097], t(89) = 0.11). These 2 monkeys were essentially identical to the other 3 monkeys whose reward schedule matched the other participants. To determine if species differences were related to other task differences, we looked at the effects of variations in the number of targets (**Experiment 2**), task timing (**Experiment 3**), and differences in response modality (**Experiment 4**) in humans.

Monkeys did a 3-target version of the task, but both mice and humans did a 2-target version. Therefore, it is possible that monkeys were more similar to humans only because adding a third target (1) improved reward acquisition, (2) reduced switching, and (3) decreased exploration. In an online sample of 150 humans (1 session each, 45 000 total trials), we manipulated the number of targets and found that the likelihood of getting rewards changed with the number of targets (**Figure 3A;** normalised difference from chance; one-way ANOVA: $F_{2, 141} = 16.03$, $p < 0.0001$, S = 144 total sessions). However, increasing the number of targets tended to increase, rather than decrease switching, though this was at trend (**Figure 3B;** one-way ANOVA: $F_{2, 141} = 2.85$, p = 0.061, S = 144 total sessions). Further, increasing the number of targets had no effect on exploration (**Figure 3C;** one-way ANOVA: $F_{2,132} = 0.75$, p = 0.48; S = 135 total sessions). This implies that changing the number of targets was not a sufficient explanation for differences in either the probability of switching or the probability of exploration between species. Manipulating the number of targets also did not alter the relative energy of exploration and exploitation (one-way ANOVA: $F_{2, 132} = 0.99$, p = 0.37; S = 135), the energy barrier between the states (one-way ANOVA: $F_{2, 132} = 2.66$, p = 0.074; S = 135), or the probability of repeating exploitation (**Figure 3D;** i.e., the transition matrix parameter; one-way ANOVA: $F_{2, 132} = 0.53$, p = 0.53; S = 135 total sessions). These results show that simply changing the number of targets cannot fully account for the species difference in persistence.

Although increasing the number of targets did not alter exploration, it did tend to increase switching (at trend). Therefore, we next asked if there were any specific changes in the probability of switching within either the exploitative, persistent regime or in the exploratory switching regime. There was no difference in switching in the persistent regime with the number of targets (One-way ANOVA: $F_{2, 139} = 1.00$, p = 0.37; S = 142 total sessions; **see Methods**), but there was a moderate increase in in switching in the exploratory regime (**Figure 3F;** One-way ANOVA: $F_{2, 139}$

= 10.25, $p < 0.0001$; S = 142 total sessions). The specific way that exploratory switching changed as a function of the number of targets implied that humans—like monkeys (Ebitz et al., 2018; Wilson et al., 2021)—used a random form of exploration in this task. Under random exploration, we would expect exploratory switching to increase systematically with the number of targets (**Figure 3E, top**; see **Methods**). This is because random choices between a smaller number of targets (i.e. 2) are more likely to repeat (i.e. 50% of the time) than random choices between a large number of targets (4 targets will repeat 25% of the time). We can directly solve for the switch probabilities expected under random sampling (see **Methods)**, and this reference line is illustrated in **Figure 3F**. By contrast, if exploratory decisions were based on the rewards of the chosen target (i.e., a win-stay, lose-shift strategy), switching would be unaffected by the number of alternatives (**Figure 3E, bottom**). We found that the human pattern of switching closely followed the prediction of a random exploration strategy (**Figure 3F, Supplementary Table 3**). In sum, while altering the number of targets did not alter persistence in humans, it did suggest that humans used a random sampling strategy for exploration.

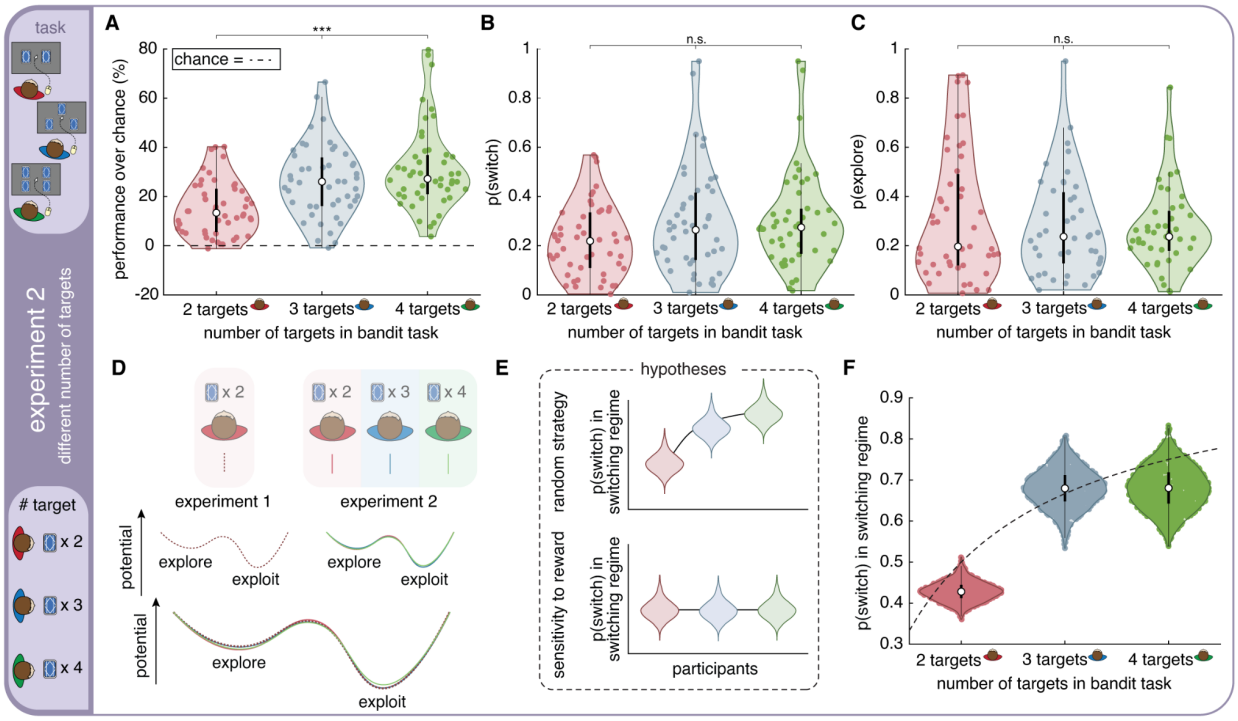


Figure 3. Effects of manipulating the number of targets and the trial length in humans performing the bandit task (Experiment 2 and 3). A) Percentage of reward relative to chance by number of targets (2, 3, or 4). **B**) Switch probability by number of targets. **C**) Probability of exploration by number of targets. **D**) State dynamic landscapes for varying numbers of targets (Same conventions as Figure 2F). **E**) Cartoon illustrating predicted relationships between the switching-regime switch probability and the number of targets under the hypothesis of random exploration (top) or reward-dependent exploration (bottom) **F**) Switch probability for the fast-switching distribution (the "switching regime") by number of targets. Thick black lines = IQR, thin = whiskers, open circle = median. Asterisks represent significance levels as follows: * indicates standard $\alpha = 0.05$ threshold; ** indicates $\alpha = 0.001$; and *** indicates $\alpha = 0.0001$.

Exploratory decision-making is affected by physiological and psychological processes that operate in the time scale of the body, not just in the time scale of trials (Shourkeshti et al., 2023). Therefore, it is possible mice were less persistent than primates because each trial took longer in this species, compared to humans and monkeys. Therefore, in **Experiment 3**, we manipulated trial lengths in humans via lengthening inter-trial interval times. In an online sample of 299 human participants (1 session each, 89,699 total trials), we found slight variations in the likelihood of getting rewards across as the inter-trial-interval increased (**Figure 4A;** normalised difference from chance; one-way ANOVA: $F_{2, 296} = 3.48$, $p < 0.05$, S = 299 total sessions). However, there was no significant effect of the inter-trial intervals on switching (**Figure 4B;** one-way ANOVA: $F_{2, 296} = 0.53$, $p = 0.59$, S = 299 total sessions) or exploration (**Figure 4C**; one-way ANOVA: $F_{2, 282} = 0.48$, $p = 0.62$; S = 285 total sessions). Again, we found no systematic differences in performance, switching or exploration by self-reported gender in this experiment (**Figure S3**). These results suggest that trial lengths did not impact the behaviours that differed between species in this task and thus is not likely to explain species differences.

Given that mice, monkeys, and humans performed bandit tasks using different modalities of response, it remained possible that the observed differences between species were simply due to differences in response modality. To determine what effect changing response modalities might have, in **Experiment 4**, we had human participants (n = 29, 1 session each) perform the task using either a touchscreen interface (14,015 total trials; similar to the mice) and an eye-tracking interface (14,485 total trials; similar to the monkeys). To ensure sufficient power, we collected enough participants to detect an effect size of 0.6—less than a quarter of the magnitude of the species-level effect on the probability of switching (Cohen's d = 2.54, between mice and monkeys) and exploration (Cohen's d = 2.42, between mice and monkeys). We found no significant differences within participants across response modalities. This was true in performance over chance (**Figure 4D;** paired t-test: $t(28) = 0.960$, $p = 0.35$), probability of switching (**Figure 4E**; $t(28) = 0.730$, $p = 0.47$), and probability of exploring (**Figure 4F**; $t(24) = -0.298$, $p = 0.77$). Thus, the effects of changing response modality are considerably smaller than the differences observed across species and are unlikely to be a sufficient explanation for them.

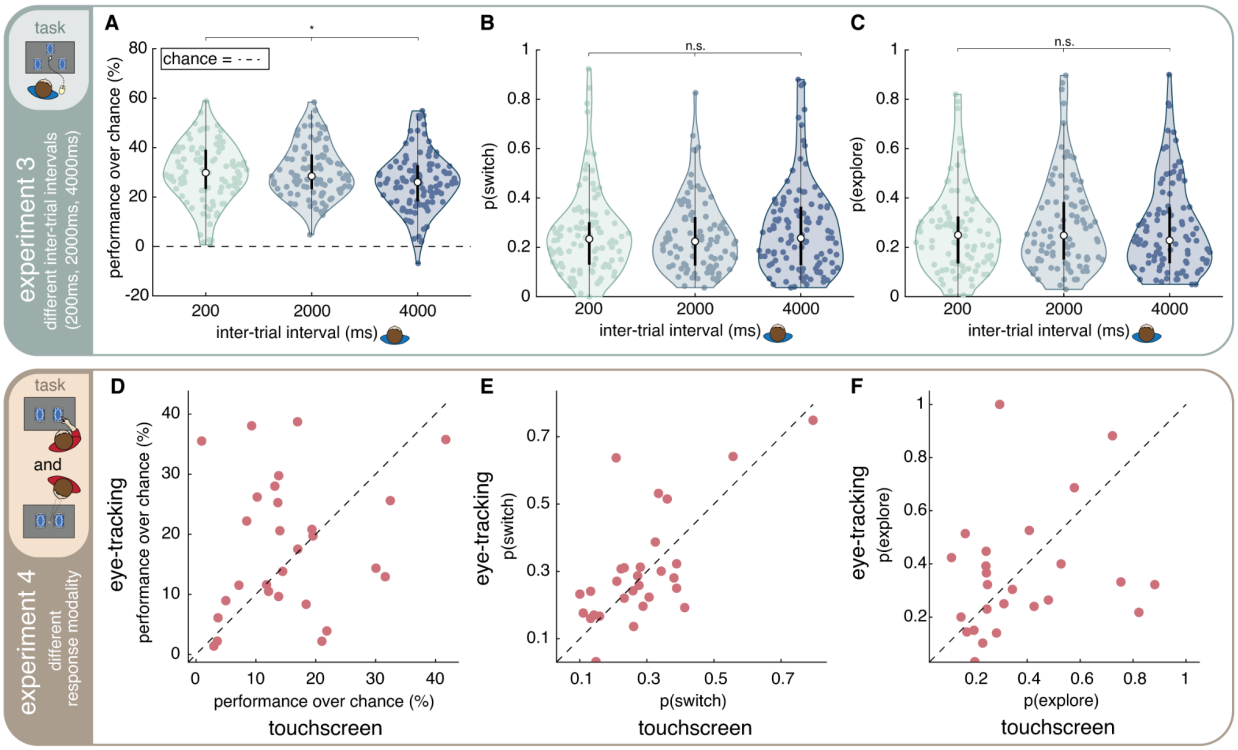


Figure 4. Effects of manipulating the trial length and modality of response in humans performing the bandit task (Experiment 3 and 4). A) Percentage of reward relative to chance across varying inter-trial interval times (200ms, 2000ms, 4000ms). Thick black lines = IQR, thin = whiskers, open circle = median. **B)** Switch probability by inter-trial interval times. **C)** Probability of exploration by inter-trial interval times. **D)** Within-participant comparison of percentage of reward relative to chance between the touchscreen and the eye tracking response modalities. **E-F)** Same as in D but for probability of switching (E), and exploration (F). Asterisks represent significance levels as follows: * indicates standard α = 0.05 threshold; ** indicates α = 0.001; and *** indicates α = 0.0001.

Learning index analysis and reward sensitivity

The primates' tendency to exploit more than mice did not appear to be an artifact of minor differences in task design or timing. Therefore, we next considered the possibility that these differences between species were due to differences in their capacity to learn from rewards. We evaluated this using a common "learning index" (a 1-trial-back measure of the effect of reward outcomes on switch decisions, normalised by the probability of switching; see **Methods**). For this analysis, we returned to the data from Experiment 1 (cross-species experiment). We found that humans appeared to learn faster compared to mice (**Figure 5A**; humans-mice: $p < 0.0001$, b = -1.43, 95% CI = [-1.86, -1.00], t(504) = -6.57), but did not differ from monkeys (humans-monkeys: $p = 0.047$, b = -1.06,95% CI = [-2.12, -0.013], t(345) = -1.99). Monkeys and mice also had a marginal difference in learning speed (**Figure 5A**; monkeys-mice: $p < 0.005$, b = -0.60, 95% CI = [-1.00, -0.21], t(341) = -2.99). However, the interpretation of this learning index is complicated because it is normalised by the overall probability of switching and primates switched less overall. This means that the learning index could change across species either because of differences in learning from reward outcomes or because of differences in switching frequently (i.e. persistence).

To differentiate between these possible explanations, we separately analysed choice patterns after rewarded or unrewarded trials. If the major difference between species was in learning from rewards, then the tendency to repeat rewarding options should positively co-vary with the tendency to switch away from non-rewarding options (**Figure 5B**, top). In short, humans should be most sensitive to reward outcomes, followed by monkeys, then mice. Conversely, if the major difference between species was in persistence, then the tendency to repeat rewarding options should be inversely related to the tendency to switch away from non-rewarding options across species (**Figure 5B**, bottom). In short, humans should be the most persistent, followed by monkeys, and then mice. Compared to mice, we found that primates persisted more with their choices, both after being rewarded (**Figure 5C**; humans-mice: $p < 0.0001$, $b = -0.18$, 95% CI = [-0.21, -0.15], $t(511) = -11{,}51$; monkeys-mice: $p < 0.0001$, $b = -0.20$, 95% CI = [-0.28, -0.12], $t(344) = -4.87$), and after not receiving a reward (**Figure 5D**; humans-mice: $p < 0.0001$, $b = 0.14$, 95% CI = [0.074, 0.20], $t(511) = 4.24$; monkeys-mice: $p < 0.0001$, $b = 0.32$, 95% CI = [0.19, 0.45], $t(344) = 4.7$). Together, these results suggest the major systematic difference between species was an increase in persistence rather than increased sensitivity to rewards, though some increase in reward sensitivity may also have been at play in humans.

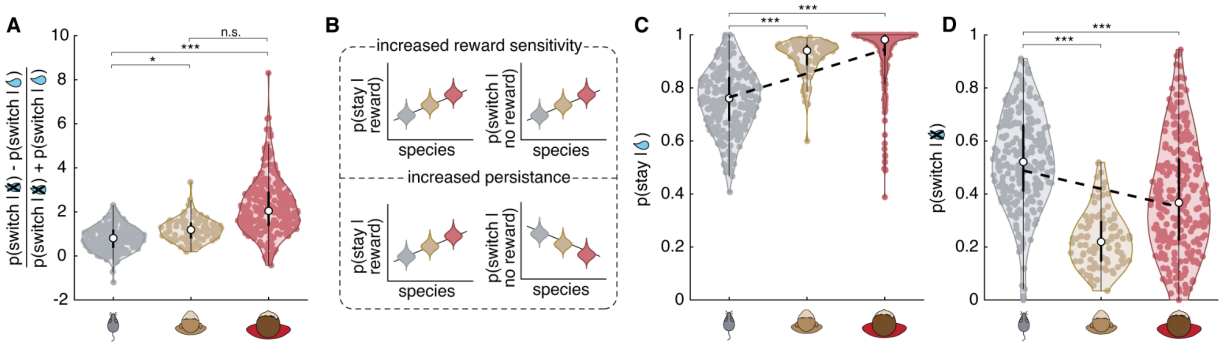


Figure 5. Learning and persistence across species A) Index of reward learning across species. Thick black lines = IQR, thin = whiskers, open circle = median. **B**) Hypothesis cartoon illustrating predictions under the hypothesis that species differences in switching were due to reward sensitivity (top) or persistence (bottom). **C**) Probability of selecting the same option after obtaining a reward, compared across species. Same conventions as A. **D**) Probability of selecting a different option after not obtaining a reward, compared across species. Same conventions as A. Asterisks represent significance levels as follows: * indicates standard $\alpha = 0.05$ threshold; ** indicates $\alpha = 0.001$; and *** indicates $\alpha = 0.0001$.

DISCUSSION

This study compared the performance and decision-making strategies of mice, monkeys, and humans in an uncertain decision-making task. All three species performed the task similarly, alternating between a strategy of rapidly switching between the options, and a strategy of persistently choosing the same option. Despite these shared strategies, we found species differences in the average performance and the tendency to switch between targets. Mice switched more frequently than primates. Computational analysis of the switching patterns revealed that the increase in switching in the mice was driven by their tendency to exploit less persistently than the primates, which meant that they initiated exploration more frequently. This

was not due to impulsivity or satiety in the mice nor was it possible to explain species differences as a function of low-level differences between the tasks like the number of options, the timing of the trials, or the modality of the response. Instead, primates, and especially humans, appeared to persist in exploiting valuable options for longer than mice did.

These results resonate with previous work looking at species differences in similar tasks. For example, in comparing mice performing a reversal learning task with monkeys performing a blocked bandit task, Woo et al., 2023 found several differences between species. Specifically, monkeys underperformed mice, switching more often than mice. Nonetheless, monkeys showed evidence of integrating past reward and choice information over longer timescales, consistent with the idea of a “choice history” component (Lau and Glimcher, 2005; Grabenhorst et al., 2019). However, it was not clear which of these prior results were due to species differences or to differences between the tasks. Because the present study harmonised the structure of the task across species, added an additional important species to the comparison (i.e., humans), and controlled for task differences via 3 additional control experiments (i.e., varying the number of targets, trial length, and response modality), the results we present here suggest a reinterpretation of that previous study. Specifically, when the tasks were equated, we found that monkeys outperformed mice and that mice switched more than monkeys. This could suggest that these specific results in Woo et al., 2023 may have been due to differences in task design. However, we do replicate the intriguing result from Woo et al., 2023 in finding evidence of a mixture of both a short and a long term tendency to switch between targets in monkeys. In our data, this longer tendency to stick with a target, which our computational approach links to exploitative states, is greater in primates compared to mice. Using this novel cross-species computational approach, we suggest that these differences may reflect species-specific variations in the stability of underlying brain states over time.

One reason why primates might persist more in their choices, could be that they had more cognitive self-control: the ability to regulate their impulses, letting them weigh long-term benefits against immediate rewards. The capacity for self-control is more prevalent in species with larger brain sizes (MacLean et al., 2014). Here, self-control could help sustain a choice policy in the absence of reward, for example, or to help animals avoid the temptation to try something new (Stillman et al., 2017). Indeed, we found that primates were more persistent with their choices and were less likely to switch from a target immediately after it did not give a reward, whereas mice did not persist as much, and quickly explored other alternatives. Thus, species differences in the capacity for self-control could help explain why both primates persisted for longer than mice did.

A second, complementary explanation for why primates persisted more than mice could be differences in neural timescales across species. Single neurons and neural populations process information with characteristic time constants, often called “neural timescales” (Zilio et al., 2021). Previous studies have shown that different brain regions have differing neural timescales (Murray et al., 2014; Golesorkhi et al., 2021; Zilio et al., 2021), perhaps tailored to the functions for each region (Hasson et al., 2008). Brain regions with longer neural timescales are better suited for integrating information over longer periods of time, like in working memory, while brain regions with shorter neural timescales are better suited for processing information that needs quick

integration, like sensory cues (Zilio et al., 2021). Notably, the prefrontal cortex (PFC), crucial for cognitive self-control (Cohen et al., 2013), cognitive functions, and decision-making (Krawczyk, 2002; Domenech and Koechlin, 2015), has been found to have longer neural timescales (Murray et al., 2014). The PFC is also more elaborated in primates compared to mice (Laubach et al., 2018; Preuss and Wise, 2022). This implies the elaborated primate PFC could improve persistence in decision-making tasks by facilitating the sustained integration of information. This contrasts with mice, whose less elaborated PFC could suggest shorter neural timescales and a reduced capacity for persistent exploitation. Of course, species differences in neural time scales could also be the underlying neural mechanism for species differences in cognitive functions, like self-control and future studies are needed to determine how individual differences in self-control and neural timescales predict differences in persistence.

There is also a third reason why primates might persist more than mice: differences in their ecological niches. Social primates, like rhesus macaques, benefit from collective vigilance within their groups (Iki and Kutsukake, 2021), this allows each individual monkey to be slightly less vigilant, and therefore lets them focus on exploiting resources for longer before looking up to scan for threats. Mice, on the other hand, are mostly prey species (Dickman, 1992) which might require them to be more vigilant and favour less sustained focus on other tasks. Differences in ecological niches across species could also result in the task being less suitable for mice as specified here. Perhaps mice are better adapted to more volatile environments, and therefore, perhaps the differences in persistence found in this task could be minimised if the task environment was more volatile.

Comparative work is essential both for understanding how the human brain evolved and for ensuring that preclinical research will translate into real-world impact in human lives. That said, comparative studies like this one also have unique challenges and real limitations. For one, whenever data is collected across multiple labs over multiple years, it introduces variability. Here, we worked to harmonise data collection across labs, but it was impossible to equate factors like training time and handling protocols. There were also differences in response and reward modalities between species, due to differences in physicality and familiarity with certain apparatuses. While none of these factors appeared to be a sufficient explanation for our major results in the sense that manipulating factors within humans could not replicate the species differences, we cannot rule out the possibility that task or training differences interacted with real species differences in complex ways. While our results are intriguing, future studies are needed to determine the effects of multiple methodological differences both between and within species. Another limitation of this study is our focus on a single stochastic decision-making task. While humans and other primates persisted more than mice in this task, future studies are still needed to determine if there are species differences in persistence in other tasks.

Multiple clinical disorders and clinically-relevant stressors affect the persistence of decision-making and executive function in humans (Verdejo-García et al., 2006; Tolin et al., 2009; Mäntylä et al., 2012; Blanco et al., 2013; Teng et al., 2016; Kaske et al., 2023). In this context, it is important to acknowledge that understanding how persistence differs across species can help advance preclinical or translational research in two major ways. It is critical that preclinical and translational studies use an animal model whose behaviour captures the aspects of cognition that

are compromised in disease. Preclinical studies in mice do not always translate well into effective human interventions (Worp et al., 2010; Perrin, 2014; Walker and Eggel, 2020). For example, despite huge investments in preclinical work on depression in mice, limitations in the behaviours typically assessed in mice as putatively linked with depression have been an impediment to the development of effective and novel treatments (National Institute of Mental Health, 2019; Planchez et al., 2019; Shemesh and Chen, 2023). If humans and mice use a different balance of strategies to navigate the world, then it is possible that interventions that change behaviour in mice would not necessarily change behaviours in humans in an exactly paralleled task. Our results suggest that humans and nonhuman primates have distinct tendencies towards persistence of decision-making, and that monkeys may be a strong animal model for disorders in which the persistence of decision-making and executive function are compromised. However, these results also provide a strong clue that if mouse tasks can be designed to maximize persistence, this may go a long way towards improving the relevance and impact of high throughput rodent preclinical testing.

METHOD

Experimental models and participant details

All animal care and experimental procedures were approved by the relevant ethical review board (**mice**: the guidelines of the National Institution of Health and the University of Minnesota; **monkeys**: the guidelines of Stanford University Institutional Animal Care and Use Committee and the Rochester University Committee on Animal Resources; **humans for Experiment 1, 3 and 4**: the guidelines of the Comité d'Éthique de la Recherche en Sciences et Santé (CERSES) of the University of Montreal; **humans for Experiment 2**: the guidelines of Princeton University Institutional Review Board). The human data and much of the monkey data has not been analysed or reported previously. Some sessions from two of the five monkeys have been analysed previously (28/58 sessions; (Ebitz et al., 2018). The mouse data has been reported previously (Chen et al., 2021b) but all analyses here are new.

All species were presented with a series of trials in which they made choices between physically identical targets that were presented in front of them on a computer screen. Specific details of each experimental setup are as follows:

Mice

Thirty-two BL6129SF1/J mice (16 males and 16 females) were obtained from Jackson Laboratories (stock #101043). Mice arrived at the lab at 7 weeks of age, and were housed in groups of four with *ad libitum* access to water and mild food restriction (85–95% of free feeding weight) for the experiment. Animals engaging in operant testing were housed in a 9AM to 9PM reversed light cycle to permit testing during the dark period. Before operant chamber training, animals were food restricted to 85–90% of free feeding body weight. Operant testing occurred five days a week (Monday-Friday). Additional information regarding mouse data collection has been reported previously (Chen et al., 2021b).

Monkeys

Five male rhesus macaques (between 5 and 15 years of age; between 6 and 16 kg) participated in this study. Three of the monkeys were singly housed and two were pair housed. All were housed in small colony rooms (6-10 animals per room). Animals were surgically prepared with head restraint prostheses before training began. Analgesics were used to minimise discomfort. After recovery, monkeys were acclimated to the laboratory and head restraint, then placed on controlled access to fluids and trained to perform the task over the course of 3 months. One animal was naive at the start of the experiment, the other four had previously participated in oculomotor and visual attention studies (2 monkeys) or decision-making studies (2 monkeys). Data was collected 5 days a week (Monday-Friday). Additional information regarding two of the monkeys has been reported previously (Ebitz et al., 2018) Data from the other 3 have not been previously analysed or reported.

Human

Humans were recruited via the online platform, Amazon Mechanical (mTurk). To avoid bots and improve data quality, participants were only accepted when they had a minimum of 5000 approved human intelligence tasks (HIT) and a minimal percentage of 98% in proportions of completed tasks that are approved by requesters. Geographical restrictions were set for US participants only. Participants were not allowed to repeat the experiment. All participants who successfully submitted the HIT were paid a base rate of $0.50 USD, plus a bonus of $3.85 mean ± $0.90 SD (for all 3 experiments, n = 707) based on their performance (for each trial that ended with a reward, participants were given a $0.02 compensation). For experiment 1, a total of 258 participants (120 females, 137 males, 1 preferred not to say) completed the task. Data was collected from 9AM to 2PM EST, to minimise data collection during night hours across coasts. For experiment 2, a total of 150 participants (gender not collected) completed the task. Data was collected from 9AM to 5PM EST. For experiment 3, a total of 299 participants (139 females, 158 males, 2 preferred not to say) completed the task. Data was collected from 9AM to 2PM EST, to minimise data collection during night hours across coasts. For experiment 4, a total of 29 participants completed the task. Data was collected from 9AM to 5PM EST. For each participant, the first response modality in which they performed—eye tracking versus touchscreen—was randomized.

Procedure

For each study, participants performed a spatial restless k-armed bandit task where they chose between physically identical visual cues. In this task, physically identical targets are presented in spatial locations that are associated with a probability of reward. Reward probabilities ranged between 0.9 and 0.1 and could diminish or increase over time at a rate that was fixed across experiments (10% chance of a step of 0.1). For 2/5 monkeys, the floor probability of reward was 0.3, rather than 0.1, to improve motivation (see **Results**).

Because rewards were variable, independent, and probabilistic, participants could only infer values through sampling the targets and integrating reward history over multiple trials. There were minor variations between the mouse, monkey, and human studies due to a combination of factors:

(1) the data was collected independently across multiple labs, (2) the tasks were adapted to the typical research approaches used in each species. For example, mice and monkeys both received a primary, liquid reinforcer as reward. Humans on the other hand received money, a secondary reinforcer. Monkeys received a 3-target version of the bandit task, mice received a 2-target version, and humans received a 2-, 3- and 4-target version. Additional variations between the tasks are described below:

Mice

Mice indicated their choices by nose-poking on a touchscreen. Each trial, mice were presented with a cue displaying two identical squares positioned on the left and right side of the screen. A nose poke to one of the target locations on the touchscreen was required to register a response. Rewards were given in the form of food pellets. Mice completed either 300 trials or spent a maximum of two hours in the operant chamber. On average, mice performed 276.38 trials (min: 46 trials, max: 300 trials) per session.

Monkeys

Monkeys indicated their choices by making saccadic eye movements towards one of three identical gabor kernels. Choices were registered when the monkeys fixated on the eccentric target for a specified minimum period (150ms). Eye position was monitored at 1000Hz via an infrared eye tracker (SR Research). Rewards were given in the form of juice. On average, monkeys performed 622.34 trials (min: 144 trials, max: 1377 trials) per session.

Humans

Humans indicated their choices by moving a computer mouse towards one of the backs of playing cards on the screen. In Experiment 1, human participants had to choose between 2 identical blue backs of playing cards. In Experiment 2, human participants had to choose between 2, 3, or 4 backs of playing cards, with each card being identical except for their colour. In Experiment 3, human participants had to choose between 3 identical blue backs of playing cards. They used a computer mouse to click the desired options and register their response. Rewards were given in the form of money ($0.02 per reward). Every human participant completed 300 trials per session, except for 1 participant who completed 299 trials during their session. Prior to the experiment, the humans completed an additional 20-25 practise trials, which were meant to familiarise them with the task but were not included in the analyses. In Experiment 4, each participant completed two successive versions of the task. In the touchscreen version, they selected targets by physically touching the display, also touching on a fixation cross located in the middle of the screen during the ITI. In the eye-tracking version, their head position was stabilized with a chin rest while an SR Research infrared eye tracker recorded their eye movements. The order of response modalities was randomized for each participant.

Quantification and Statistical Analysis

General analysis techniques

Data was analysed with custom MATLAB scripts and p-values were compared against the standard $\alpha = 0.05$ threshold for species-specific paired t-tests, and one-way ANOVAs. When conducting the three planned pairwise species comparisons for Experiment 1 (humans vs. mice, monkeys vs. mice, and humans vs. monkeys), we applied a Bonferroni correction to control the family-wise error rate ($\alpha' = 0.05/3 \approx 0.017$). We analysed the data using a general linear mixed-effects model, with Species as a fixed factor and subjects nested within species as a random factor to account for repeated measures on the same animal. In cases where we compared mice vs monkeys separately, we included sessions as a fixed factor (plus its interaction with species), however, while these were included in the model, we did not report the session and interaction effect in the paper. The Session term wasn't included for humans, since each human completed only a single session. In **Experiment 1**, the sample size for mice (n = 32; total sessions = 255), for monkeys n = 5, total sessions = 93, for humans (2-armed bandit task) n = 258, total sessions = 258. In **Experiment 2**, the sample size for humans was (2-armed bandit task) n = 50, total sessions = 50, for humans (3-armed bandit task) n = 50, total sessions = 50, for humans (4-armed bandit task) n = 50, total sessions = 50. In **Experiment 3**, the sample size for humans was (ITI 200ms) n = 99, total sessions = 99, for humans (ITI 2000ms) n = 93, total sessions = 93, for humans (ITI 4000ms) n = 107, total sessions = 107. In **Experiment 4**, the sample size for humans was (2-armed touchscreen and eye-tracking bandit task) n = 29, total sessions = 29.

A small number of sessions from some participants were excluded from analysis. Specifically, 6 sessions out of 150 were excluded in **Experiment 2** because participants did not select all available targets during the session and this experiment specifically looked at the behavioural effects of manipulating the number of targets. Otherwise, sessions were only excluded from specific analyses when these analyses were impossible, given the participants' behaviour. For example, in **Experiment 1**, 7 sessions out of 606 (3 for mice, 0 for monkeys, 4 for humans) were excluded from certain analyses of switching behaviour (i.e. the learning index, mixture model, HMM) because the participant did not switch during those sessions. However, these sessions were included in all other analyses, including of the probability of switching (i.e. **Figure 1D**). The specific exclusion criteria for each analysis as well as the number of excluded sessions is described within the relevant section of the Methods. For response time analyses, any trial exceeding 1.5 times the interquartile range (IQR) was excluded as an outlier. Response time analyses excluded 14 sessions from 93 monkeys because no reaction time data could be recovered from these sessions.

To analyse changes in behaviour over time within sessions, we compared the first 150 and last 150 trials because most sessions had at least these many trials (**Figure S2A**). In the rare cases where sessions contained fewer than 300 trials (48/255 sessions in mice, 14/93 sessions in monkeys, none in humans), data at the extremities was replicated in both bins. Since mouse response time data had a heavy right tail, the plot in **Figure S2A** (mice) is not centered at 1.

Random Exploration Among k-Arms

In a random exploration strategy, a target is selected at random on each trial. This means that the probability of repeating a choice is the same as the independent probability of making that

choice (i.e. it is always 1/k, where k is the number of options). The probability of switching away from a previous option is then the probability of choosing any other option:

$$p(\text{switch}) \;=\; \frac{k-1}{k} \tag{1}$$

Note that as k increases, as the number of targets increases, the probability of switching increases systematically, under the hypothesis that decisions are made randomly. This is the explicit equation, with 0 free parameters, that is plotted in **Figure 3F**.

Exponential Mixture Distribution

We analysed the temporal structure of the participants' choice sequences with a mixture model. If a single time constant (probability of switching) governed the behaviour, we would expect to see exponentially distributed inter-switch intervals. That is, the distribution of inter-switch intervals should be well described by the following model:

$$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \tag{2}$$

Where β is the "survival parameter" of the model: the average inter-switch interval. However, although the time between switch decisions was largely monotonically decreasing and concave upwards, the distribution was not well described by a single exponential distribution (**Figure 2A**). Participants had more short-latency and more long latency choice runs, indicating that a single switching probability could not have generated the data. Therefore, we fit mixtures of varying numbers of exponential distributions (1-4) to all species (**Figure 2A**), in order to infer the number of switching regimes in these choice processes. For continuous-time processes, these mixture distributions would be of the form:

$$f(x) = \sum_{i=1}^{n} \pi_i e^{-\frac{x}{\beta_i}} \tag{3}$$

Where $1 \geq \pi_i \geq 0$ for all π_i, and $\Sigma_i \pi i = 1$. Here, each β_i reflects the survival parameter (average inter-switch interval) for each component distribution i and the π_i reflects the relative weight of each component. Because trials were discrete, we fit the discrete analog of this distribution: mixtures of 1-4 discrete exponential (geometric) distributions (Barger, 2006). Mixtures were fit via the expectation-maximisation algorithm and we used standard model comparison (Burnham and Anderson, 2002) to determine the most probable number of mixing components (**Figure 2A, Results**).

We used a bootstrap procedure to illustrate the distribution of mixture model parameters in **Figure 2B** and **Figure 3F**. This meant that we resampled, with replacement, from the sessions collected

in each species to generate bootstrapped distributions of run lengths (n distributions = 1000, number of sessions equal to the data). We then fit the exponential mixture model to each sample of run lengths, giving a bootstrapped estimate of mixture model parameters. (N.B. Statistical analyses were done on the raw, non-bootstrapped, data, the bootstrapping was only done for illustration.)

Some participants had to be excluded from mixture model analyses because their distribution of run lengths prevented the identification of model parameters. This could happen either because they either had fewer than 2 switches between options (i.e. it was impossible to measure any run lengths) or because their run lengths lacked the variation required for the expectation maximisation algorithm to function (i.e. all run lengths were identical). In **Experiment 1**, 4 human sessions (of 606) were excluded (11 total, when including the 7 excluded previously because no switches were observed) in the main results. In **Figure S2B** of the supplementary materials, sessions were divided into their first 150 and last 150 trials, and any session with no switches in either half was excluded—removing an additional 6 mouse, 6 monkey, and 39 human sessions. In **Experiment 2**, 2 sessions out of 150 were excluded (8 total, including the 6 excluded previously for not choosing all available targets).

Hidden Markov Model (HMM)

In order to identify how often different species were exploring or exploiting, we fit an HMM to each session of each species. Here, choices (y) are “emissions” that are generated by an unobserved decision process that is in some latent, hidden state (z). Latent states are defined by both the probability of making each choice k (out of N_k possible options), and by the probability of transitioning from each state to every other state. Our model consisted of two types of states, the explore state and the exploit state. The emissions model for the explore state was uniform across the options:

$$p(y_t = k | z_t = explore) = \frac{1}{N_k} \tag{4}$$

This is the maximum entropy distribution for a categorical variable—the distribution that makes the fewest number of assumptions about the true distribution and thus does not bias the model towards or away from any particular type of high-entropy choice period. This does not require, imply, impose, or exclude that decision-making happening under exploration is random (Ebitz et al., 2019, 2020). Because exploitation involves repeated sampling of each option, exploit states only permitted choice emissions that matched one option. That is:

$$\begin{cases} p(y_t = k | z_t = exploit_i, k \in exploit_i) = 1 \\ p(y_t = k | z_t = exploit_i, k \notin exploit_i) = 0 \end{cases} \tag{5}$$

The latent states in this model are Markovian, meaning that they are time-independent. They depend only on the most recent state (z_t):

$$p(z_t|z_{t-1}, y_{t-1}, \dots, z_1, y_1) = p(z_t|z_{t-1}) \quad (6)$$

This means that we can describe the entire pattern of dynamics in terms of a single transition matrix. This matrix is a system of stochastic equations describing the one-time- step probability of transitioning between every combination of past and future states (i, j).

$$p(z_t = i|z_{t-1} = j) \quad (7)$$

Because the task varied in the number of available targets, the number of states scaled accordingly. In the version with two targets, participants could occupy three states (one explore state and two exploit states; i.e. one for each target). In the three-target version, they had four states (one explore state and three exploit states). And in the four-target version, they had five states in total (one explore state alongside four exploit states). To produce long, exponentially-distributed runs of repeated choices to a single target, the HMM had one latent exploitative state for each target. To produce short, random run lengths, the HMM had one shared explore state from which decisions to any of the choices were equally likely. For all participants, parameters were tied across exploit states such that each exploit state had the same probability of beginning (from exploring) and of sustaining itself. Transitions out of the exploration, into exploitative states, were similarly tied. The model also assumed that the participants had to pass through exploration in order to start exploiting a new option, even if only for a single trial. This is because the utility of exploration is to maximise information about the environment (Mehlhorn et al., 2015). If a participant switches from a bout of exploiting one option to another option, that very first trial after switching should be exploratory because the outcome or reward contingency of that new option is unknown and that behaviour of switching aims to gain information. Through fixing the emissions model, constraining the structure of the transition matrix, and tying the parameters, the final HMM had only two free parameters: one corresponding to the probability of exploring, given exploration on the last trial, and one corresponding to the probability of exploiting, given exploitation on the last trial.

The model was fit via expectation-maximisation using the Baum Welch algorithm (Bilmes, 2000). This algorithm finds a (possibly local) maxima of the complete-data likelihood. A complete set of parameters θ includes the emission and transition models, discussed already, but also initial distribution over states. Because the participants had no knowledge of the environment at the first trial of the session, we assumed they began by exploring, rather than adding another parameter to the model here. The algorithm was reinitialized with random seeds 20 times, and the model that maximised the observed (incomplete) data log likelihood across all the sessions for each animal was ultimately taken as the best. To ultimately infer latent states from choices, we used the forward-backward algorithm to label the most probable sequence of latent states.

Some participants were excluded from analyses that depended on the HMM because the model did not fit these participants. This totalled 59 sessions out of 1084 (>5.5%, 17 for mice, 0 for monkeys, 15 for humans in **Experiment 1**, 9 in **Experiment 2**, 14 in **Experiment 3**, and 4 in **Experiment 4**). The HMM model could fail to fit for 2 reasons: (1) because participants only chose a single target for the whole session (making model parameters unidentifiable) or (2) because fitting procedure resulted in a solution that violated the assumption of longer choice runs under

exploitation compared to exploration (where the probability of stopping a bout of exploitation was an obvious outlier in the distribution of this parameter across all species; threshold for exclusion set at 0.4).

Analysing the modulated indices:

In order to determine if response time would reveal species differences in task strategy, we looked at how response time changed as a function of certain task events. Because species performed slightly different variations of the task, with differing response modalities (i.e. touch screen for mice, saccades for monkeys, and computer mouse for humans), we were unable to compare raw response times directly. Instead, we computed "modulated indices of reaction times" by normalising the averaged difference in response times between 2 conditions by the standard deviation of response time. We specifically assessed whether response time was modulated by exploration (versus exploitation) and by reward (versus no reward) in any species. If response time was faster in exploration trials, as opposed to exploitation trials, in some species, the modulated index would be above 0, implying that exploration is "impulsive" (faster). If response time is slower for exploration, the modulated index would be below 0, implying that exploration is a slower, more "deliberate" choice process relative to exploitation (**Figure 2G** - top). A similar approach was applied to compare rewarded and unrewarded trials (**Figure 2G** - bottom). The explore/exploit modulation index could only be calculated in sessions where (1) response time data was available, and (2) that included both exploration and exploitation (560 total sessions / 606 sessions). Analyses of the rewarded/unrewarded modulation index included all sessions where response time data was available (592 total sessions / 606 sessions).

Analysing HMM Dynamics (State Dynamic Landscapes)

In order to understand the dynamics of exploration and exploitation, we analysed the HMMs. Here, we use the term "dynamics" to mean the equations that govern how a system evolves over time. In fitting our HMMs, we were fitting a set of equations that describe these dynamics: the probability of transitions between exploration and exploitation and vice versa. To illustrate how goal dynamics differed across groups, we performed certain thermodynamic analyses of the long-term behaviour of the fitted equations, generating insight into the potential energy of each state in each species (**Figure 2C**).

In statistical mechanics, processes within a system (like a decision-maker at some moment in time) occupy states (like exploration or exploitation). States have energy associated with them, related to the long-time scale probability of observing a process in those states. A low-energy state is one that is very stable and deep, much like a valley between two mountain peaks. Low-energy states will be over-represented in the system's long-term behaviour. A high energy state, like the top of a mountain, is less stable. High-energy states will be under-represented in the system's behaviour. The probability of observing a process in a given state i is related to the energy of that state (E_i) via the Boltzman distribution:

$$p_i = \frac{1}{Z} e^{\frac{-E_i}{k_B T}} \tag{8}$$

where Z is the partition function of the system, k_B is the Boltzman constant, and T is the temperature. If we focus on the ratio between two state probabilities, the partition functions cancel out and the relative occupancy of the two states is now a function of the difference in energy between them:

$$\frac{p_i}{p_j} = e^{\frac{-\left(E_i - E_j\right)}{k_B t}} \tag{9}$$

Rearranging, we express the difference in energy between two states as a function of the difference in the long-term probability of those states being occupied:

$$ln\left(\frac{p_i}{p_j}\right) k_B T = E_j - E_i \tag{10}$$

Meaning that the difference in the energetic depth of the states (the Gibbs Free Energy) is proportional to the natural log of the probability of each state, up to some multiplicative factor $k_B T$. To calculate the probability of exploration and exploitation (p_i and p_j), we solved for the stationary distribution of the fitted HMMs. The stationary distribution is the equilibrium probability distribution over states. This means that this distribution is the relative frequency of each state that we would observe if the model's dynamics were run for an infinite period of time. Each entry of the model's transition matrix reflects the probability that the participant would move from one state (e.g. exploring) to another (e.g. exploiting) at each moment in time. Because the parameters for all the exploitation states were tied, each transition matrix effectively had two states—an explore state and a generic exploit that described the dynamics of all exploit states. Each of the k sessions had its own transition matrix (A_k), which describes how the entire system—an entire probability distribution over states—would evolve from time point to time point. We observe how the dynamics evolve any probability distribution over states (π) by applying the dynamics to this distribution:

$$\pi_{t+1} = \pi_t A_k \tag{11}$$

Over many time steps, ergodic systems reach a point where the state distributions are unchanged by continued application of the transition matrix as the distribution of states reaches its equilibrium. That is, in stationary systems, there exists a stationary distribution, π, such that:

$$\pi = \pi A_k \tag{12}$$

If it exists, this distribution is a (normalised) left eigenvector of the transition matrix A_k with an eigenvalue of 1, so we solved for this eigenvector to determine the stationary distribution of each A_k. We then took an average of these stationary distributions across all sessions for each group, and plugged these back into the Boltzman equations to calculate the relative energy (depth) of exploration and exploitation as illustrated in **Figure 2F** and **Figure 3D**.

In order to understand the dynamics of exploration and exploitation, we need to not only understand the depth of the two states, but also the height of the energetic barrier between them: the energy required to transition from exploration to exploitation and back again. Here, we build on the Arrhenius equation from chemical kinetics that relates the rate of transitions (k) between some pair of states to the activation energy required to affect these transitions (E_a):

$$k = Ae^{\frac{E_a}{k_B T}} \tag{13}$$

where A is a constant pre-exponential factor related to the readiness of reactants to undergo the transformation. We set this to one. Again, k_BT is the product of temperature and the Boltzman constant. Note the similarities between this equation and the Boltzman distribution illustrated earlier. Rearranging to solve for activation energy yields:

$$E_a = -ln\left(\frac{k}{A}\right) k_B T \tag{14}$$

Thus, activation energy, much like the relative depth of each state, is also proportional to some measurable function of behaviour, up to some multiplicative factor k_BT. Note that our approach has only identified the energy of three discrete states (an explore state, an exploit state, and the peak of the barrier between them). These are illustrated by tracing a continuous potential through these three points to provide a physical intuition for the differences in explore/exploit dynamics between species

To create the attractor basin graphs, transition matrices were calculated individually for all participants (Seed = 20), and then averaged across groups, see Methods section: Analysing HMM Dynamics (state dynamic landscapes) for more details. All statistical tests used and statistical details were reported in the results.

Acknowledgments

The authors would like to thank Devin Heinz-Kehoe, Mariem Zid, and Rishabh Singhal for invaluable discussion, A. David Redish for feedback on a prior version of this manuscript, and Bastian Bechtold for developing software for the violin plots in the results.

Funding:
NSERC Discovery Grant RGPIN-2020-05577 (RBE)
Research Corporation for Science Advancement (RBE)
Frederick Gardner Cottrell Foundation Project #29087 (RBE)
Canada Research Chair Dynamics of Cognition FD507106 (RBE)
Research Fellowship from the Jacobs Foundation (RBE)
Unifying Neuroscience and Artificial Intelligence – Québec UNIQUE (VJL)
National Institutes of Health grants R21 MH127607 and K23 MH050909 (ABH)
National Institutes of Health grants P50 MH119569 and R01 MH123661 (NMG)

Author contributions:
Conceptualization: VJL, RBE
Methodology: RBE, NMG, ABH
Investigation: RBE, CSC, AS, ABH
Visualization: VJL
Supervision: RBE, NMG, ABH
Writing—original draft: VJL, RBE
Writing—review & editing: VJL, RBE, CSC, NMG, ABH

Competing interests: Authors declare that they have no competing interests.

Data and Code Availability: Data and software are available upon request to the Lead Contact (Becket Ebitz, rebitz@gmail.com)

Figures and Tables

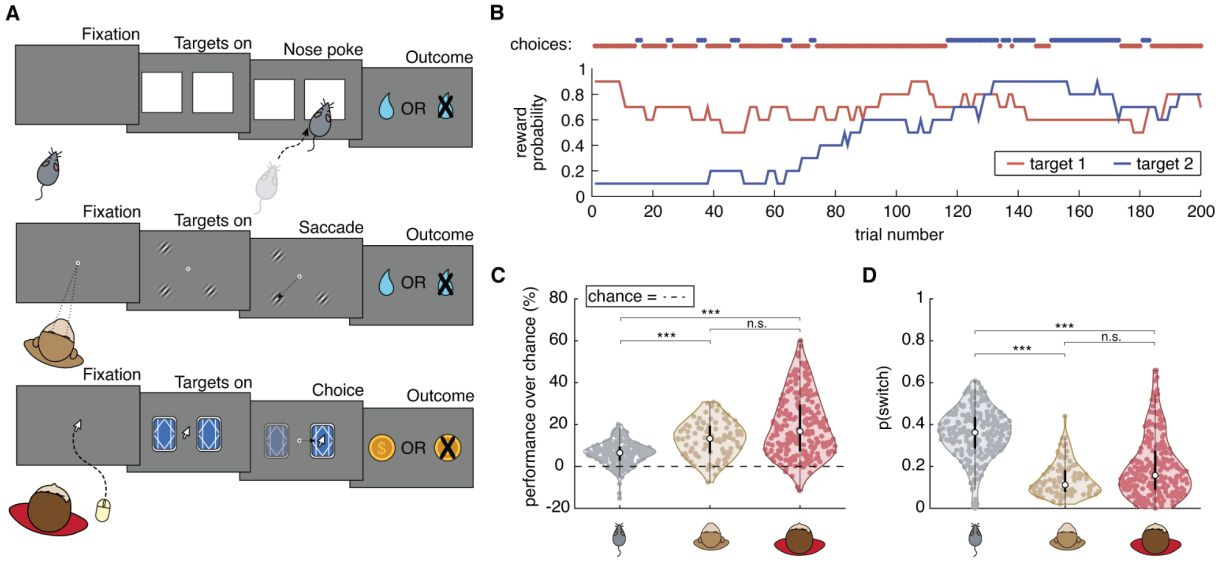


Figure 1. Task design and behaviour across species. **A**) A schematic representation of the bandit task in each species (mice = top, monkeys = middle, humans = bottom). **B**) Example reward schedule, including 200 trials from one session with one human. The reward probabilities of each of the 2 targets (blue and red traces) walk randomly, independently across trials. The humans' choices are illustrated as coloured dots along the top. **C**) Percentage of reward relative to chance in all species. Thick black lines = IQR, thin = whiskers, open circle = median. Black dotted line = chance performance. **D**) Probability of switching targets during the task between species. Same conventions as C. Asterisks represent significance levels as follows: * indicates standard α = 0.05 threshold; ** indicates α = 0.001; and *** indicates α = 0.0001, corrected for multiple comparisons.

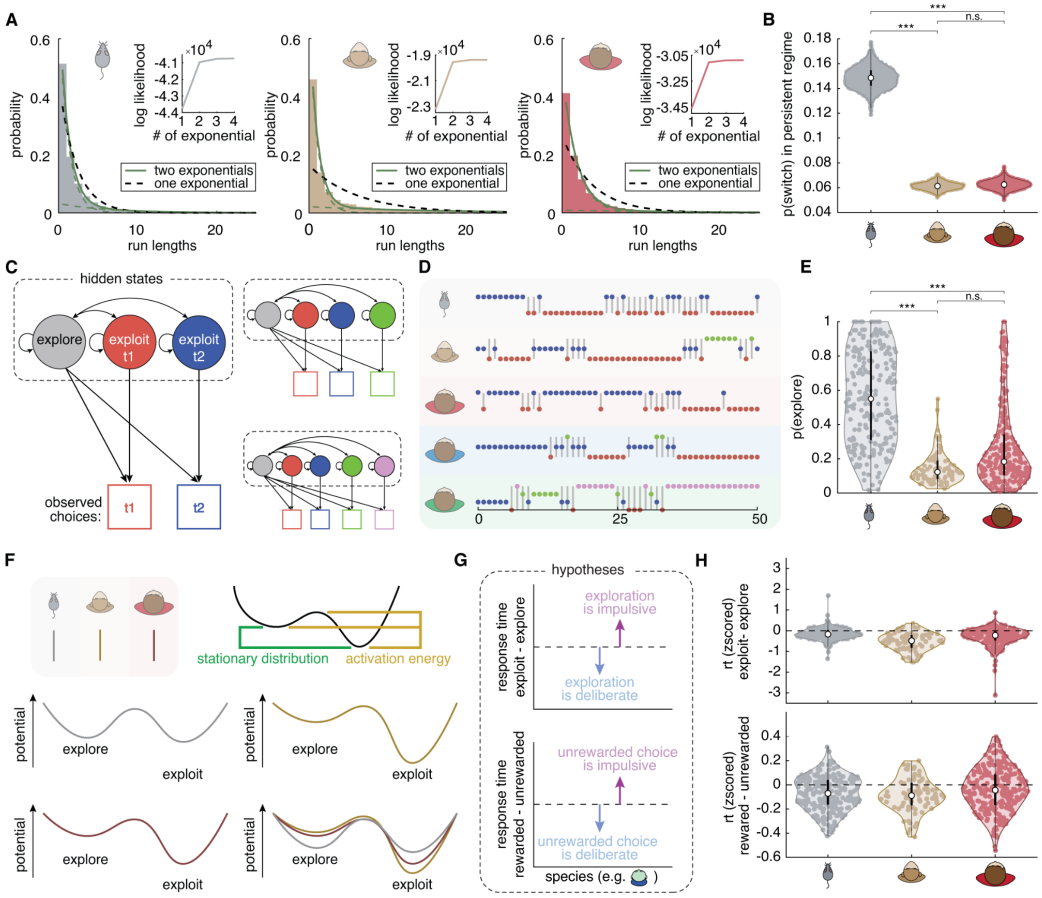


Figure 2. Different patterns of switching and exploration across species. A) Distributions of the number of trials between switch decisions ("run lengths") in mice, monkeys and humans. If the species had a fixed probability of switching, run lengths would be exponentially distributed (black dotted line). A mixture of two exponential distributions (purple line) suggests 2 distinct probabilities of switching. Dotted purple lines show each mixing distribution, one slow-switching and another fast-switching. (Inset) Log likelihoods for different mixture models containing a range of 1 to 4 exponential distributions in each species. **B**) Bootstrapped estimates of switch probability for the slow-switching distribution (the "persistent regime") across species. Thick black lines = IQR, thin = whiskers, open circle = median. **C**) Hidden Markov models (HMMs) were used to infer the goal state on each trial from the sequence of choices. The model included one persistent state for each target (''exploit'') and one state in which subjects could choose any of the targets ("explore"). Right) The model can be extended to account for different numbers of targets by adding exploit states. **D**) Fifty-trial example choice sequences for mice, monkeys and humans. The coloured circles represent the chosen target, and the grey lines highlight the explore choices identified with the HMM. **E**) Probability of exploration across species, as inferred by the HMM. Same conventions as B. **F**) Fitting the HMM involves identifying a set of equations that describe the dynamics of exploration and exploitation, meaning the rate at which participants explore, exploit, and switch between states. Left) Certain analytic measures of these equations, namely their stationary distributions (Boltzmann, 1868) and activation energies (Arrhenius, 1889) can be used to derive an intuitive picture of the landscape of state dynamics. Middle) Average state dynamic landscapes for each species. Right) State dynamic landscapes for all species overlaid G) Differences in response time (RT) can reveal which choices are fast and impulsive or slow deliberative. Top: a cartoon illustrating RT differences under the hypothesis that exploration is more deliberative than exploitation (blue) or that exploration is more impulsive (purple). Bottom: same for rewarded and unrewarded trials. H) Z-scored RT differences across all three species. Top: difference across HMM-defined exploration and exploitation. Bottom: between rewarded and unrewarded trials. Thick black lines = IQR, white circle = median. Asterisks represent significance levels as follows: * indicates standard α = 0.05 threshold; ** indicates α = 0.001; and *** indicates α = 0.0001, corrected for multiple comparisons.

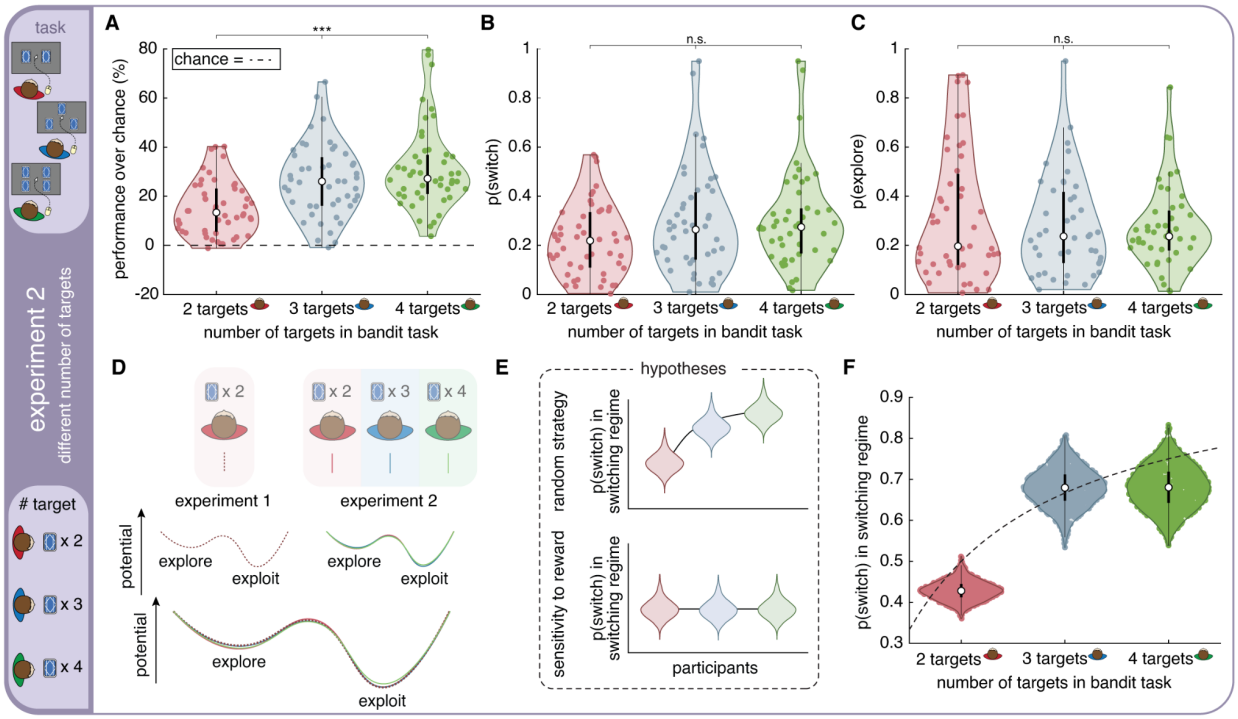


Figure 3. Effects of manipulating the number of targets and the trial length in humans performing the bandit task (Experiment 2 and 3). A) Percentage of reward relative to chance by number of targets (2, 3, or 4). **B**) Switch probability by number of targets. **C**) Probability of exploration by number of targets. **D**) State dynamic landscapes for varying numbers of targets (Same conventions as Figure 2F). **E**) Cartoon illustrating predicted relationships between the switching-regime switch probability and the number of targets under the hypothesis of random exploration (top) or reward-dependent exploration (bottom) **F**) Switch probability for the fast-switching distribution (the "switching regime") by number of targets. Thick black lines = IQR, thin = whiskers, open circle = median. Asterisks represent significance levels as follows: * indicates standard α = 0.05 threshold; ** indicates α = 0.001; and *** indicates α = 0.0001.

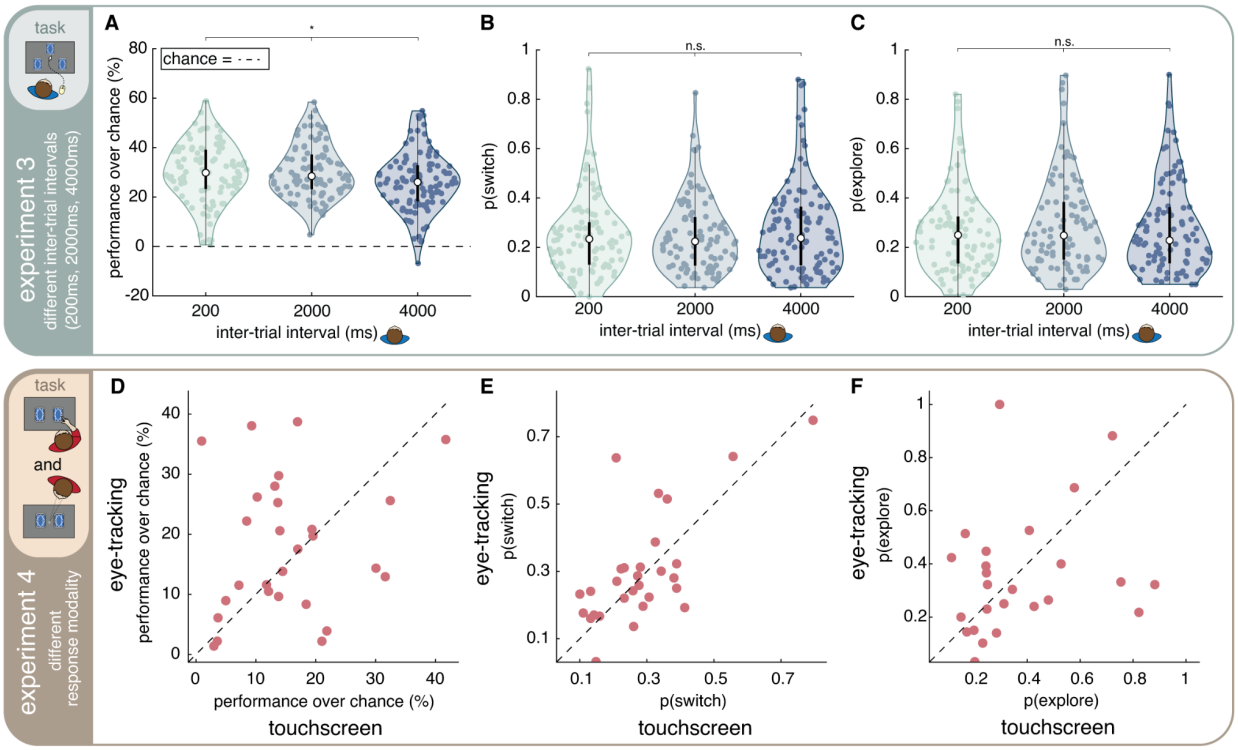
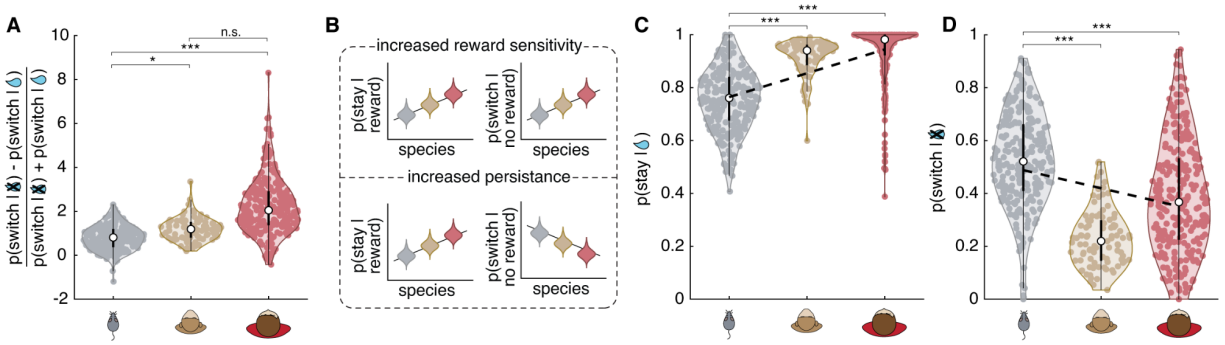


Figure 4. Effects of manipulating the trial length and modality of response in humans performing the bandit task (Experiment 3 and 4). A) Percentage of reward relative to chance across varying inter-trial interval times (200ms, 2000ms, 4000ms). Thick black lines = IQR, thin = whiskers, open circle = median. **B)** Switch probability by inter-trial interval times. **C)** Probability of exploration by inter-trial interval times. **D)** Within-participant comparison of percentage of reward relative to chance between the touchscreen and the eye tracking response modalities. **E-F)** Same as in D but for probability of switching (E), and exploration (F). Asterisks represent significance levels as follows: * indicates standard α = 0.05 threshold; ** indicates α = 0.001; and *** indicates α = 0.0001.



964
965 **Figure 5. Learning and persistence across species A**) Index of reward learning across species. Thick black lines =
966 IQR, thin = whiskers, open circle = median. **B**) Hypothesis cartoon illustrating predictions under the hypothesis that
967 species differences in switching were due to reward sensitivity (top) or persistence (bottom). **C**) Probability of selecting
968 the same option after obtaining a reward, compared across species. Same conventions as A. **D**) Probability of selecting
969 a different option after not obtaining a reward, compared across species. Same conventions as A. Asterisks represent
970 significance levels as follows: * indicates standard α = 0.05 threshold; ** indicates α = 0.001; and *** indicates α =
971 0.0001.
972

REFERENCES

Arrhenius S (1889) Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. Z Für Phys Chem 4U:226–248.

Barger KJ-A (2006) Mixtures of Exponential Distributions to Describe the Distribution of Poisson Means in Estimating the Number of Unobserved Classes. Available at: https://ecommons.cornell.edu/handle/1813/2953 [Accessed June 16, 2022].

Bari BA, Grossman CD, Lubin EE, Rajagopalan AE, Cressy JI, Cohen JY (2019) Stable Representations of Decision Variables for Flexible Behavior. Neuron 103:922-933.e7.

Bilmes J (2000) A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Tech Rep ICSI-TR-97-021 Univ Berkeley 4.

Blanco NJ, Otto AR, Maddox WT, Beevers CG, Love BC (2013) The influence of depression symptoms on exploratory decision-making. Cognition 129:563–568.

Boltzmann L (1868) Studien uber das gleichgewicht der lebenden kraft. Wissenschafiliche Abh 1:49–96.

Boyden ES, Zhang F, Bamberg E, Nagel G, Deisseroth K (2005) Millisecond-timescale, genetically targeted optical control of neural activity. Nat Neurosci 8:1263–1268.

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd ed. New York: Springer.

Chen CS, Ebitz RB, Bindas SR, Redish AD, Hayden BY, Grissom NM (2021a) Divergent Strategies for Learning in Males and Females. Curr Biol CB 31:39-50.e4.

Chen CS, Knep E, Han A, Ebitz RB, Grissom NM (2021b) Sex differences in learning from exploration Izquierdo A, Wassum KM, Izquierdo A, eds. eLife 10:e69748.

Cohen JR, Berkman ET, Lieberman MD (2013) Intentional and Incidental Self-Control in Ventrolateral Prefrontal Cortex. In: Principles of Frontal Lobe Function (Tranel D, Stuss DT, Knight RT, eds), pp 0. Oxford University Press. Available at: https://doi.org/10.1093/med/9780199837755.003.0030 [Accessed April 19, 2024].

Dickman CR (1992) Predation and Habitat Shift in the House Mouse, Mus Domesticus. Ecology 73:313–322.

Disotell TR, Tosi AJ (2007) The monkey’s perspective. Genome Biol 8:226.

Domenech P, Koechlin E (2015) Executive control and decision-making in the prefrontal cortex. Curr Opin Behav Sci 1:101–106.

Ebitz RB, Albarran E, Moore T (2018) Exploration Disrupts Choice-Predictive Signals and Alters Dynamics in Prefrontal Cortex. Neuron 97:450-461.e9.

1007 Ebitz RB, Sleezer BJ, Jedema HP, Bradberry CW, Hayden BY (2019) Tonic exploration governs
1008 both flexibility and lapses. PLoS Comput Biol 15:e1007475.

1009 Ebitz RB, Tu JC, Hayden BY (2020) Rules warp feature encoding in decision-making circuits.
1010 PLOS Biol 18:e3000951.

1011 Ellenbroek B, Youn J (2016) Rodent models in neuroscience research: is it a rat race? Dis
1012 Model Mech 9:1079–1087.

1013 Ernst PB, Carvunis A-R (2018) Of mice, men and immunity: a case for evolutionary systems
1014 biology. Nat Immunol 19:421–425.

1015 Gibbs RA et al. (2007) Evolutionary and biomedical insights from the rhesus macaque genome.
1016 Science 316:222–234.

1017 Golesorkhi M, Gomez-Pilar J, Zilio F, Berberian N, Wolff A, Yagoub MCE, Northoff G (2021)
1018 The brain and its time: intrinsic neural timescales are key for input processing. Commun
1019 Biol 4:1–16.

1020 Grabenhorst F, Tsutsui K-I, Kobayashi S, Schultz W (2019) Primate prefrontal neurons signal
1021 economic risk derived from the statistics of recent reward experience Lee D, Ivry RB,
1022 Louie K, eds. eLife 8:e44838.

1023 Groman SM, Smith NJ, Petrullli JR, Massi B, Chen L, Ropchan J, Huang Y, Lee D, Morris ED,
1024 Taylor JR (2016) Dopamine D3 Receptor Availability Is Associated with Inflexible
1025 Decision Making. J Neurosci Off J Soc Neurosci 36:6732–6741.

1026 Grossman CD, Bari BA, Cohen JY (2022) Serotonin neurons modulate learning rate through
1027 uncertainty. Curr Biol 32:586-599.e7.

1028 Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A Hierarchy of Temporal Receptive
1029 Windows in Human Cortex. J Neurosci 28:2539–2550.

1030 Iki S, Kutsukake N (2021) Japanese macaques relax vigilance when surrounded by kin. Anim
1031 Behav 179:173–181.

1032 Iyer ES, Weinberg A, Bagot RC (2022) Ambiguity and conflict: Dissecting uncertainty in
1033 decision-making. Behav Neurosci 136:1–12.

1034 Izquierdo A, Aguirre C, Hart EE, Stolyarova A (2019) Rodent Models of Adaptive Value
1035 Learning and Decision-Making. Methods Mol Biol Clifton NJ 2011:105–119.

1036 Kaske EA, Chen CS, Meyer C, Yang F, Ebitz B, Grissom N, Kapoor A, Darrow DP, Herman AB
1037 (2023) Prolonged physiological stress is associated with a lower rate of exploratory
1038 learning that is compounded by depression. Biol Psychiatry Cogn Neurosci
1039 Neuroimaging 8:703–711.

1040 Knox W, Otto A, Stone P, Love B (2012) The Nature of Belief-Directed Exploratory Choice in
1041 Human Decision-Making. Front Psychol 2 Available at:
1042 https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00398 [Accessed October 24,
1043 2023].

1044 Krawczyk DC (2002) Contributions of the prefrontal cortex to the neural basis of human decision
1045 making. Neurosci Biobehav Rev 26:631–664.

1046 Lau B, Glimcher PW (2005) Dynamic Response-by-Response Models of Matching Behavior in
1047 Rhesus Monkeys. J Exp Anal Behav 84:555–579.

1048 Laubach M, Amarante LM, Swanson K, White SR (2018) What, If Anything, Is Rodent Prefrontal
1049 Cortex? eNeuro 5 Available at: https://www.eneuro.org/content/5/5/ENEURO.0315-
1050 18.2018 [Accessed April 19, 2024].

1051 MacLean EL et al. (2014) The evolution of self-control. Proc Natl Acad Sci 111:E2140–E2148.

1052 Manger P, Cort J, Ebrahim N, Goodman A, Henning J, Karolia M, Rodrigues S-L, Strkalj G
1053 (2008) Is 21st century neuroscience too focussed on the rat/mouse model of brain
1054 function and dysfunction? Front Neuroanat 2 Available at:
1055 https://www.frontiersin.org/articles/10.3389/neuro.05.005.2008 [Accessed October 18,
1056 2023].

1057 Mäntylä T, Still J, Gullberg S, Del Missier F (2012) Decision Making in Adults With ADHD. J
1058 Atten Disord 16:164–173.

1059 Mehlhorn K, Newell BR, Todd PM, Lee MD, Morgan K, Braithwaite VA, Hausmann D, Fiedler K,
1060 Gonzalez C (2015) Unpacking the exploration–exploitation tradeoff: A synthesis of
1061 human and animal literatures. Decision 2:191–215.

1062 Murray JD, Bernacchia A, Freedman DJ, Romo R, Wallis JD, Cai X, Padoa-Schioppa C,
1063 Pasternak T, Seo H, Lee D, Wang X-J (2014) A hierarchy of intrinsic timescales across
1064 primate cortex. Nat Neurosci 17:1661–1663.

1065 National Institute of Mental Health (2019) NOT-MH-19-053: Notice of NIMHs Considerations
1066 Regarding the Use of Animal Neurobehavioral Approaches in Basic and Pre-clinical
1067 Studies. Available at: https://grants.nih.gov/grants/guide/notice-files/NOT-MH-19-
1068 053.html [Accessed May 26, 2025].

1069 Perrin S (2014) Preclinical research: Make mouse studies work. Nature 507:423–425.

1070 Planchez B, Surget A, Belzung C (2019) Animal models of major depression: drawbacks and
1071 challenges. J Neural Transm Vienna Austria 1996 126:1383–1408.

1072 Preuss TM, Wise SP (2022) Evolution of prefrontal cortex. Neuropsychopharmacol Off Publ Am
1073 Coll Neuropsychopharmacol 47:3–19.

1074 Rich AS, Gureckis TM (2018) Exploratory choice reflects the future value of information.
1075 Decision 5:177–192.

1076 Saddoris MP, Sugam JA, Stuber GD, Witten IB, Deisseroth K, Carelli RM (2015) Mesolimbic
1077 Dopamine Dynamically Tracks, and Is Causally Linked to, Discrete Aspects of Value-
1078 Based Decision Making. Biol Psychiatry 77:903–911.

1079 Shemesh Y, Chen A (2023) A paradigm shift in translational psychiatry through rodent
1080 neuroethology. Mol Psychiatry 28:993–1003.

1081 Shourkeshti A, Marrocco G, Jurewicz K, Moore T, Ebitz RB (2023) Pupil size predicts the onset
1082 of exploration in brain and behavior. BioRxiv Prepr Serv Biol:2023.05.24.541981.

1083 Siegal ML, Bergman A (2002) Waddington's canalization revisited: Developmental stability and
1084 evolution. Proc Natl Acad Sci 99:10528–10532.

1085 Soltani A, Izquierdo A (2019) Adaptive learning under expected and unexpected uncertainty.
1086 Nat Rev Neurosci 20:635–644.

1087 Stevenson TJ, Alward BA, Ebling FJP, Fernald RD, Kelly A, Ophir AG (2018) The Value of
1088 Comparative Animal Research: Krogh's Principle Facilitates Scientific Discoveries.
1089 Policy Insights Behav Brain Sci 5:118.

1090 Stillman PE, Medvedev D, Ferguson MJ (2017) Resisting Temptation: Tracking How Self-
1091 Control Conflicts Are Successfully Resolved in Real Time. Psychol Sci 28:1240–1258.

1092 Teng C, Otero M, Geraci M, Blair RJR, Pine DS, Grillon C, Blair KS (2016) Abnormal decision-
1093 making in generalized anxiety disorder: Aversion of risk or stimulus-reinforcement
1094 impairment? Psychiatry Res 237:351–356.

1095 Tolin DF, Kiehl KA, Worhunsky P, Book GA, Maltby N (2009) An exploratory study of the neural
1096 mechanisms of decision making in compulsive hoarding. Psychol Med 39:325–336.

1097 Trepka E, Spitmaan M, Bari BA, Costa VD, Cohen JY, Soltani A (2021) Entropy-based metrics
1098 for predicting choice behavior based on local response to reward. Nat Commun 12:6567.

1099 Verdejo-García A, Pérez-García M, Bechara A (2006) Emotion, Decision-Making and
1100 Substance Dependence: A Somatic-Marker Model of Addiction. Curr Neuropharmacol
1101 4:17–31.

1102 Waddington CH (1942) Canalization of Development and the Inheritance of Acquired
1103 Characters. Nature 150:563–565.

1104 Walker RL, Eggel M (2020) From Mice to Monkeys? Beyond Orthodox Approaches to the Ethics
1105 of Animal Model Choice. Animals 10:77.

1106 Wilson RC, Bonawitz E, Costa VD, Ebitz RB (2021) Balancing exploration and exploitation with
1107 information and randomization. Curr Opin Behav Sci 38:49–56.

1108 Woo JH, Aguirre CG, Bari BA, Tsutsui K-I, Grabenhorst F, Cohen JY, Schultz W, Izquierdo A,
1109 Soltani A (2023) Mechanisms of adjustments to different types of uncertainty in the
1110 reward environment across mice and monkeys. Cogn Affect Behav Neurosci 23:600–
1111 619.

1112 Worp HB van der, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, Macleod MR (2010)
1113 Can Animal Models of Disease Reliably Inform Human Studies? PLOS Med
1114 7:e1000245.

1115 Yan X, Ebitz RB, Grissom N, Darrow DP, Herman AB (2023) A low dimensional manifold of
1116 human exploratory behavior reveals opposing roles for apathy and anxiety. BioRxiv
1117 Prepr Serv Biol:2023.06.19.545645.

1118 Zilio F, Gomez-Pilar J, Cao S, Zhang J, Zang D, Qi Z, Tan J, Hiromi T, Wu X, Fogel S, Huang Z,
1119 Hohmann MR, Fomina T, Synofzik M, Grosse-Wentrup M, Owen AM, Northoff G (2021)
1120 Are intrinsic neural timescales related to sensory processing? Evidence from abnormal
1121 behavioral states. NeuroImage 226:117579.