## Title

# Persistent Decision-Making in Mice, Monkeys, and Humans

## Authors

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

## Affiliations

[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:
R. Becket Ebitz*[1]
*Corresponding Author: rebitz@gmail.com

## Abstract

Humans have the capacity to persist in behavioural policies, even in challenging environments that lack immediate reward. Our 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 an uncertain 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 much longer than the mice. We speculate that the similarities in persistence patterns in humans and monkeys, as opposed to mice, may be linked to ecological, neurobiological, or cognitive factors that differ systematically between these species.

## Teaser

Humans, monkeys and mice use similar decision-making strategies, but exploit valuable options for different lengths of time.

## MAIN TEXT

## 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.

The need for comparative analyses of human and non-human exploratory decision-making is especially urgent because 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 there has been an increasing use of mice for 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 a recent 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). Although these studies have led to fundamental insights, the overarching goal in both psychology and neuroscience remains understanding human cognition and diseases, by translating findings from animal studies into applications in humans. Achieving this ultimate goal requires 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 related 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 new alternative options? Although all three species exhibited distinct signatures of exploration and exploitation, there were large differences in how often 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 did. Control analyses and 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 an increased capacity to persist in exploitative states. If this is the case, then it may be this capacity for persistence that is perturbed by environmental and/or neurobiological challenges.

## Results

In **Experiment 1**, mice (N=32, 8 sessions each, total of S = 256 sessions, 70 778 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**). There were some minor differences between species in the timing of the task and the number of targets (**see Methods**), which we controlled for via 2 additional experiments in humans (**Experiment 2** and **Experiment 3**).

There were differences in performance measures between mice, monkeys, and humans. The species differed in the likelihood of getting rewards (normalised difference from chance; **Figure 1C**; 3-way ANOVA: $F_{2,\,311} = 265.95$, $p < 0.0001$, S = 607 total sessions), with humans performing better than monkeys who performed better than mice. There were also species differences in the probability of switching between the targets (**Figure 1D**; 3-way ANOVA: $F_{2,\,311} = 353.64$, $p < 0.0001$, S = 607 total sessions), with primates switching less often than mice (3-way ANOVA: $F_{1,\,315} = 370.81$, $p < 0.0001$, S = 607 total sessions).

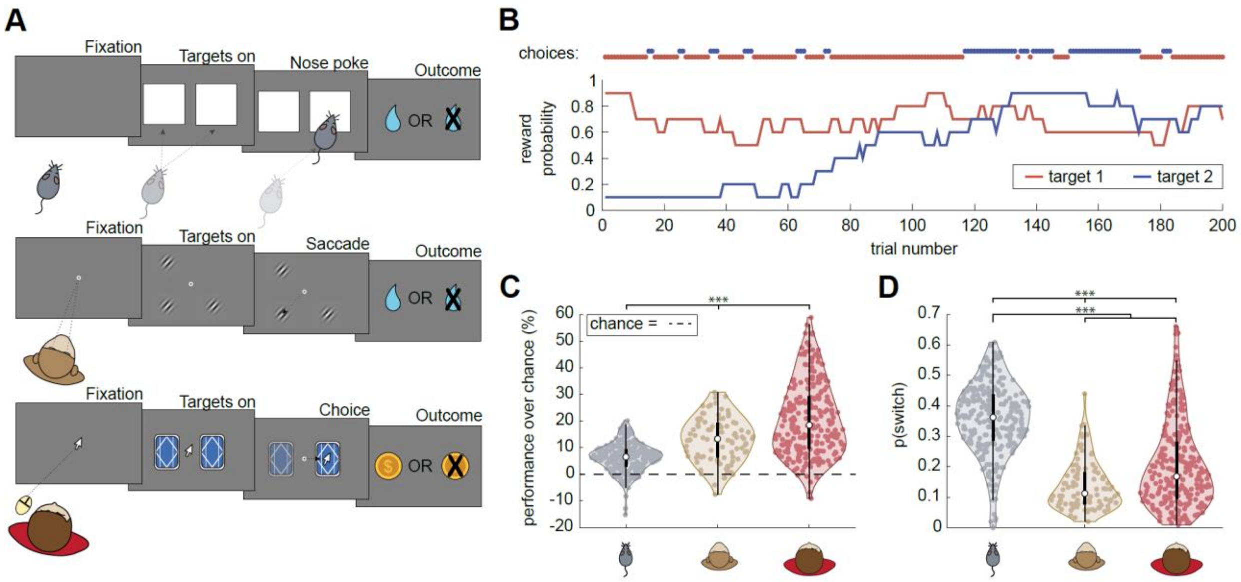

**Figure 1. Task design and behaviour across species**. **A**) A schematic representation of the bandit task in each species (mice = top, monkeys = middle, humans = bottom). **B**) Example reward schedule, including 200 trials from one session with one human. The reward probabilities of each of the 2 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. In figure, asterisks represent significance levels as follows: * indicates p < 0.05, ** indicates p < 0.001, and *** indicates p < 0.0001.

## 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 the types of switching behaviours differed across all species, we fit a "mixture model" to the distribution of interswitch intervals (number of trials between switches) in each species (**Figure 2A**; **see Methods** for more details; (Ebitz et al., 2018; Chen et al., 2021b)).

We found that the behaviour of all species could be best described as a mixture of two modes (**Figure 2A, Table S1**). Participants sometimes switched between targets at a fast pace ("switching regime") and they sometimes stuck to choosing one target repeatedly ("persistent regime"). The species differed in their (1) average switching probability during the persistent regime (3-way ANOVA: $F_{2,\,308} = 85.6$, $p < 0.0001$, S = 596 total sessions), (2) the average switching probability during the switching regime (3-way ANOVA: $F_{2,\,308} = 50.79$, $p < 0.0001$, S = 596 total sessions), and (3) the relative frequency of both regimes (3-way ANOVA: $F_{2,\,308} = 4.66$, $p < 0.02$; **Table S2**), with primates switching less often, and therefore being more persistent with their goals while in the persistent regime (3-way ANOVA: $F_{1,\,312} = 82.44$, $p < 0.0001$; **Figure 2B,** S = 596 total sessions). Monkeys and humans did not differ in their probability of switching in the persistent regime (3-way ANOVA: $F_{1,\,87} = 0.74$, $p > 0.39$, S = 343 total sessions). 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 across all species (**Figure 2E**; 3-way ANOVA: $F_{2,\,284}$ = 212.72, p < 0.0001, S = 567 total sessions). The difference between mice and primates explained the most variance between the groups (primates vs mice, $\eta^2$ = 0.1766, 17.66% of the variance; 3-way ANOVA: $F_{1,\,288}$ = 233.7, p < 0.0001, S = 567 total sessions; monkeys vs humans, $\eta^2$ = 0.0372, 3.72% of the variance, 3-way ANOVA: $\mathrm{F}_{1,\,87}$ = 71.71, p < 0.0001, S = 338 total sessions). 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 and dynamics 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: 3-way ANOVA: $F_{2,\,287}$ = 78.77, p < 0.0001, S = 562 total sessions), where mice were less likely to stay in exploitation than either primate (mice: 0.78 ± 0.13 STD across sessions, monkeys: 0.95 ± 0.03, humans: 0.87 ± 0.13). In analysing model dynamics, we considered (1) the difference in potential energy between exploration and exploitation (see Methods), and (2) the activation energy needed to transition from exploitation to exploration (**Figure 2F**). In mice, we found that exploration and exploitation had roughly the same level of stability (mean difference in energy = 0.16 ± 1.45 STD across sessions), whereas exploitation tended to be a deeper, more energetically stable state than exploration in both primates (monkeys: -2.15 ± 0.77; humans: -1.35 ± 1.79; sig. differences across species, 3-way ANOVA: $\mathrm{F}_{2,\,282}$ = 186.1, p < 0.0001, S = 558 total sessions). The amount of energy required to end a bout of exploration also differed between species: less energy was required to start to explore in the mouse compared to the monkeys and humans (differences in activation energy; 3-way ANOVA: $\mathrm{F}_{2,\,282}$ = 66.85, p < 0.0001, S = 558 total sessions; mice = 1.80 ± 1.28 STD, monkeys = 3.18 ± 0.52, humans = 2.61 ± 1.71 STD). In short, primates had a deeper, more energetically stable kind of exploitation than mice, suggesting that the differences we observed in switching behaviour and exploration could be due to the fact that primates are capable of persisting in their exploitative policies for longer than mice.

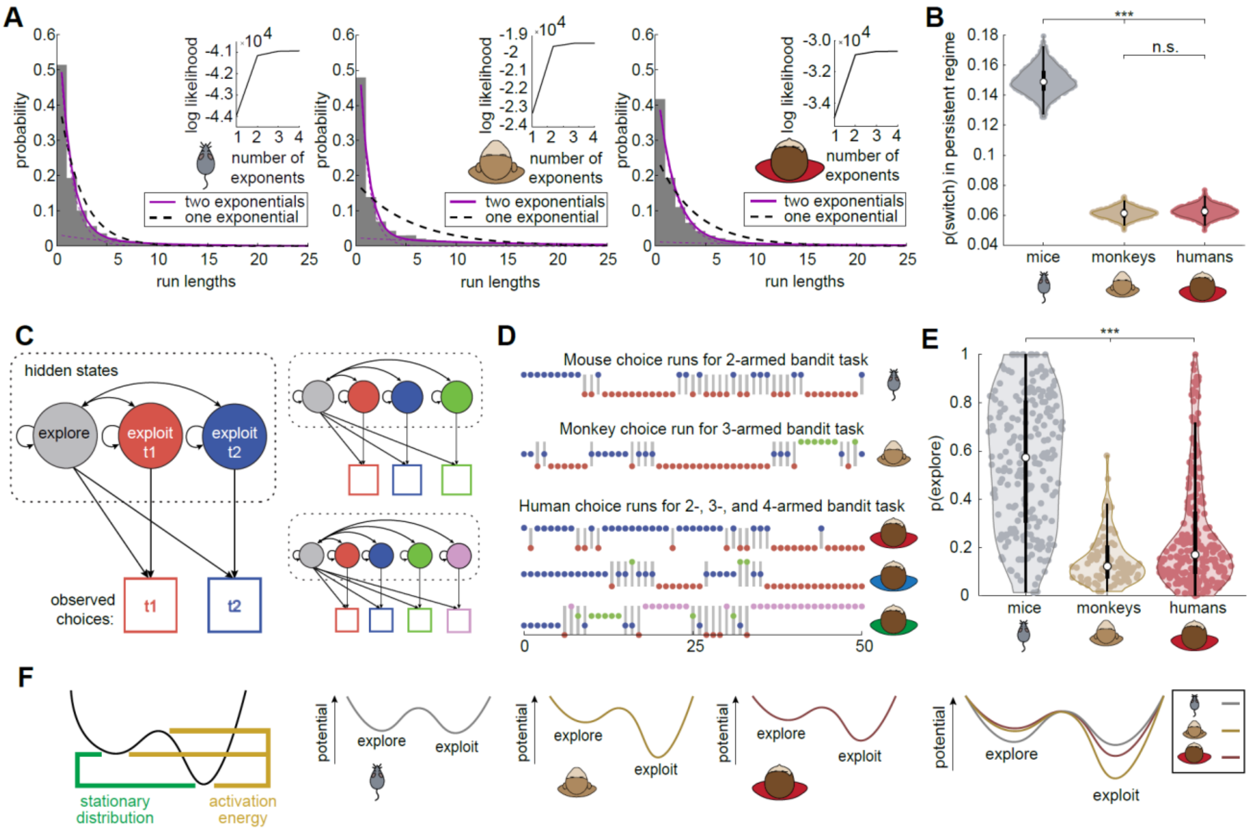

**Figure 2. Different patterns of switching and exploration across species. A**) Distributions of the number of trials between switch decisions ("run lengths") in mice, monkeys and humans. If the species had a fixed probability of switching, run lengths would be exponentially distributed (black dotted 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. In figure, asterisks represent significance levels as follows: * indicates $p < 0.05$, ** indicates $p < 0.001$, and *** indicates $p < 0.0001$.

## ***Manipulating task variables to understand species differences***

Putative species differences in rewards and persistence could be artifacts of 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. These 2 monkeys also switched less than the 3 monkeys whose reward schedule matched the other participants (3-way ANOVA: $F_{1,90} = 6.59$, $p < 0.02$, S = 93 total sessions). However, we found that excluding these 2 monkeys from the analyses did not alter the major results.

Mice still switched more than either primate species (comparing mice to all primates (all humans and all monkeys; 3-way ANOVA: $F_{1,315} = 370.81$, $p < 0.0001$, $\eta^2 = 0.1670$, 16.70%, S = 607 total sessions; comparing mice to all humans and 3 monkeys with same reward walks: 3-way ANOVA: $F_{1,280} = 333.32$, $p < 0.0001$, $\eta^2 = 0.1665$, 16.65%, S = 572 total sessions). Mice also explored more (comparing mice to all human and all monkeys: 3-way ANOVA: $F_{1,288} = 233.7$, $p < 0.0001$, $\eta^2 = 0.1766$, 17.66%, S = 567 total sessions; comparing mice to all humans and 3 monkeys with same reward walks: 3-way ANOVA: $F_{1,253} = 196.69$, $p < 0.0001$, $\eta^2 = 0.1734$, 17.34%, S = 532 total sessions).

To control for other potential confounds, we looked at the effects of variations in the number of targets (**Experiment 2**) and task timing (**Experiment 3**) 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 variations in the likelihood of getting rewards across the number of targets (normalised difference from chance; 2-way ANOVA: $F_{2,\,143} = 15.82$, $p < 0.0001$, S = 144 total sessions; **Figure 3A**). However, the effect of increasing the number of arms had only a trend-level effect on switching (2-way ANOVA: $F_{2,\,143} = 2.79$, $p = 0.065$, S = 144 total sessions; **Figure 3B**) and no effect on exploration (2-way ANOVA: $F_{2,135} = 0.59$, $p = 0.586$; S = 136 total sessions; **Figure 3C**). Thus, differences in the number of targets is not likely to explain differences in persistence between species.

Manipulating the number of targets did suggest that humans may, like monkeys (Ebitz et al., 2018; Wilson et al., 2021), use random strategies for exploration in this task. If humans were exploring randomly, we would expect the rate of switching during exploratory switching regime to vary systematically with the number of targets. 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). Thus, random exploration predicts a specific upward trend in the rate of switching with the number of targets (**Figure 3D, top**; see **Methods**). Many other types of exploration would produce no trend in switch probability as a function of the number of targets, however. If exploration focused only on the rewards of the chosen target, for example, switching would be unaffected by the number of alternatives (**Figure 3D, bottom**). We found that humans switched more frequently as more targets were available (2-way ANOVA: $F_{2,\,138} = 10.2$, $p < 0.0001$; S = 142 total sessions; see **Methods**, **Figure 3E, Table S3**). Critically, the pattern of switching closely followed the prediction from the random exploration strategy, calculated directly with 0 free parameters (see **Methods**). Manipulating the number of targets also increased the probability of staying in exploitation (2-way ANOVA: $F_{2,\,133} = 8.13$, $p < 0.0005$; S = 137 total sessions; **Figure 3F**), though it did not alter the relative energy of exploration and exploitation (2-way ANOVA: $F_{2,\,133} = 0.80$, $p = 0.45$; S = 137) or the energy barrier between states (2-way ANOVA: $F_{2,\,133} = 0.87$, $p = 0.42$; S = 137). Nonetheless, differences in the number of targets could therefore at least partly explain why monkeys had a deeper exploitation basin compared to humans and mice.

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 the inter-trial-interval (normalised difference from chance; 2-way ANOVA: $F_{2,\,295}$ = 3.04, p < 0.05, S = 299 total sessions; **Figure 3G**). However, there was no significant effect of the inter-trial interval times on switching (2-way ANOVA: $F_{2,\,295}$ = 0.38, p = 0.685, S = 299 total sessions; **Figure 3H**) or exploration (2-way ANOVA: $F_{2,\,285}$ = 0.16, p = 0.849; S = 289 total sessions **Figure 3I**). 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.

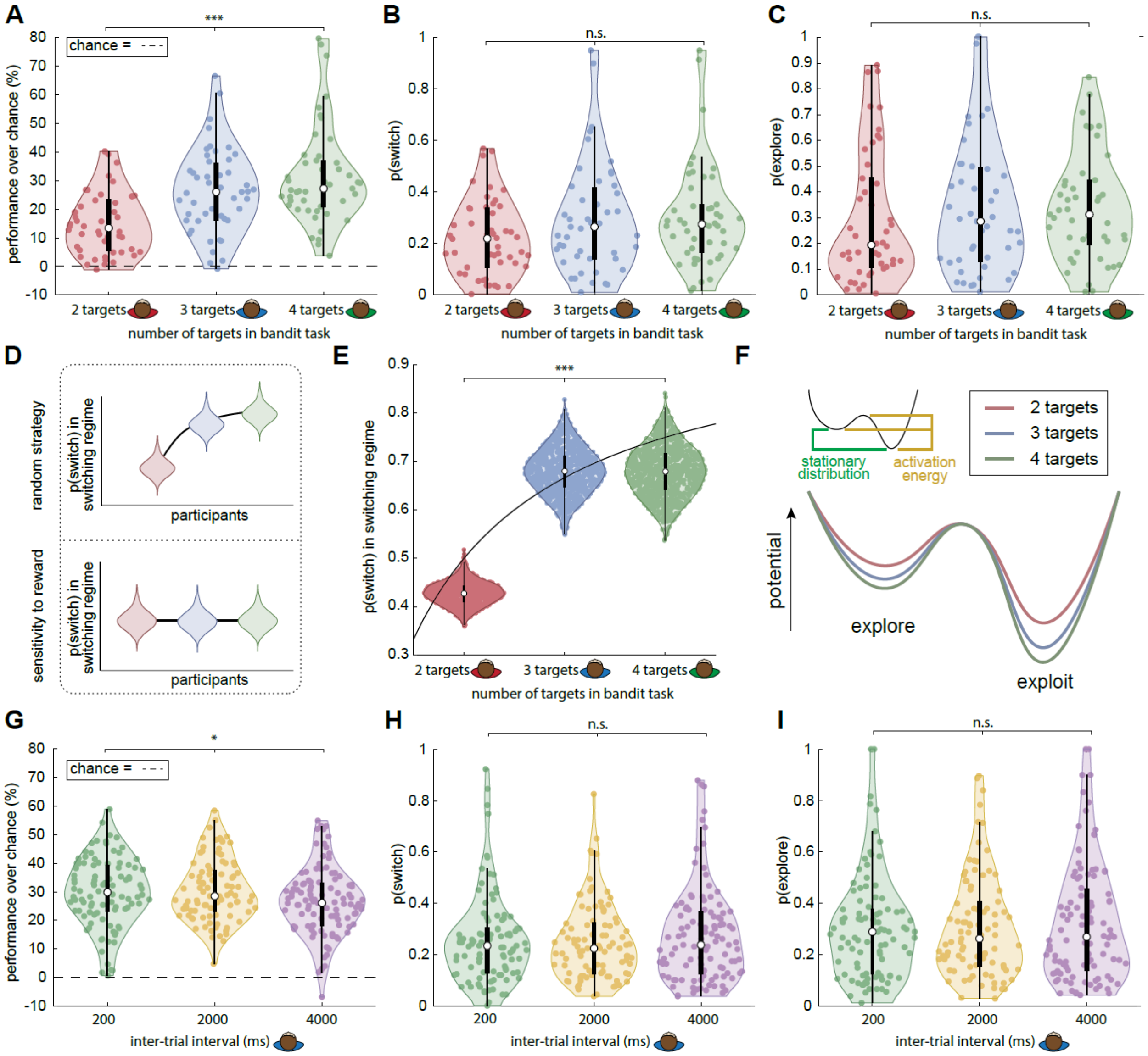

**Figure 3. Effects of manipulating the number of targets and the trial length in humans performing the bandit task (Experiment 2 and 3). A)** Percentage of reward relative to chance by number of targets (2, 3,

or 4). Thick black lines = IQR, thin = whiskers, open circle = median. **B**) Switch probability by number of targets. **C**) Probability of exploration by number of targets. **D**) Cartoon illustrating predicted relationships between the switching-regime switch probability and the number of arms under the hypothesis of random exploration (top) or reward-dependent exploration (bottom). **E**) Switch probability for the fast-switching distribution (the "switching regime") by number of targets. **F**) State dynamic landscapes for varying numbers of targets (Same conventions as Figure 2F). **G-I**) Same as A-C across varying inter-trial interval times (200ms, 2000ms, 4000ms). Thick black lines = IQR, thin = whiskers, open circle = median. In figure, asterisks represent significance levels as follows: * indicates $p < 0.05$, ** indicates $p < 0.001$, and *** indicates $p < 0.0001$.

## ***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 variations across species (**Figure 4A**; 3-way ANOVA: $F_{2,308} = 575.44$, $p < 0.0001$, S = 600 total sessions), with humans appearing to learn the fastest, then monkeys, then mice. 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 4B**, 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 4B**, 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 4C**; 3-way ANOVA: $F_{1,311} = 323.34$, $p < 0.0001$, S = 607 total sessions), and after not receiving a reward (**Figure 4D**; 3-way ANOVA: $F_{1,\,311} = 211.84$, $p < 0.0001$, S = 607 total sessions). 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.

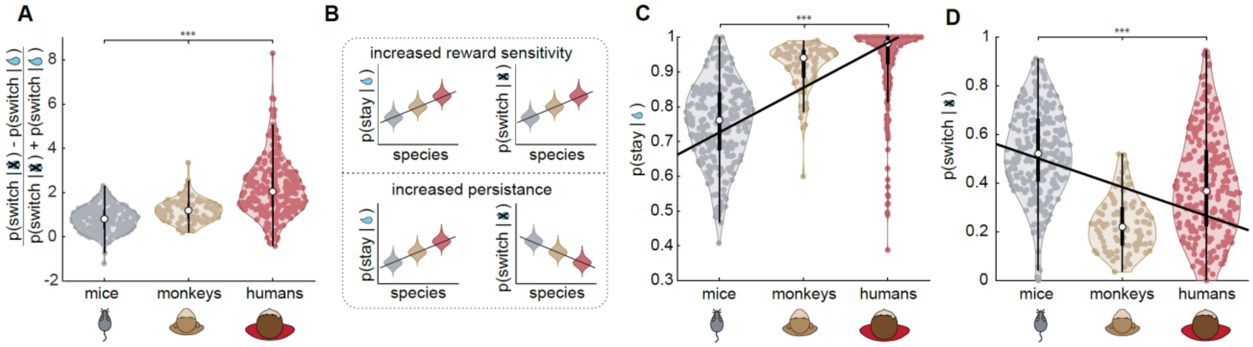

**Figure 4. Learning and persistence across species A**) Index of reward learning across species. Thick black lines = IQR, thin = whiskers, open circle = median. **B**) Hypothesis cartoon illustrating predictions under the hypothesis that species differences in switching were due to reward sensitivity (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. In figure, asterisks represent significance levels as follows: * indicates p < 0.05, ** indicates p < 0.001, and *** indicates p < 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 explore more frequently, compared to primates. Species differences were not due to low level differences between the tasks like the number of options or the timing of the trials. Instead, primates, and especially humans, appeared to persist in exploiting valuable options for longer than the mice did.

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 in their choices and were able to resist switching options immediately in the absence of a reward, while mice lost interest more quickly. 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.

Ultimately, comparative work is essential both for understanding how the human brain evolved and for ensuring that preclinical research can translate into real-world impact in human lives. Comparative studies also have unique challenges. Whenever data is collected across multiple labs over multiple years, it introduces variability. Species also necessarily differ in factors like training time and researchers tend to use different response and reward modalities in different 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, we cannot rule out the possibility that task or training differences interacted with real species differences in complex ways. We made efforts to harmonise the datasets across species, include important control experiments and analyses, and to transparently describe the methodological differences between the tasks, but our results remain suggestive. We found that humans and other primates persisted more than mice in a stochastic decision-making task, but future studies are still needed to determine if species differences in persistence are apparent in other tasks and whether these species differences can be modulated by altering certain aspects of task design. This is especially important because preclinical studies in mice do not always translate well into clinically relevant human interventions (Worp et al., 2010; Perrin, 2014; Walker and Eggel, 2020). If our results are correct, they suggest that monkeys could and should be used as a vital step in cross-species translation, particularly in the domains of decision-making and executive function.

## Materials and Methods

### ***Experimental Design***

For each study, participants performed a spatial restless k-armed bandit task. 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. Although these 2 animals switched slightly less frequently than the other 3 animals (11.89% vs 17.14%), excluding these animals from the analyses did not change any major results (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 display with two identical squares. 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 276.50 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.

## ***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 and 3**: 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.

## ***Statistical Analysis***

Data was analysed with custom MATLAB scripts and p-values were compared against the standard α = .05 threshold. 3-way ANOVAs were used to determine decision-making differences across species, unless otherwise specified. The ANOVAs modelled session-averaged data and included main effects of species, individuals (nested within species) and session number (nested within species and individuals). To minimise redundancy, only the main effect of species was reported in the paper. In experiment 1, the sample size for mice was n = 32, 256 total sessions, for monkeys n = 5, 93 total sessions, for humans (2-armed bandit task) n = 258, 258 total sessions. In experiment 2, the sample size for humans was (2-armed bandit task) n = 50, 50 total sessions, for humans (3-armed bandit task) n = 50, 50 total sessions, for humans (4-armed bandit task) n = 50, 50 total sessions. In experiment 3, the sample size for humans was (ITI 200ms) n = 99, 99 total sessions, for humans (ITI 2000ms) n = 93, 93 total sessions, for humans (ITI 4000ms) n = 107, 107 total sessions.

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 607 (3 for mice, 0 for monkeys, 4 for humans) were excluded from certain analyses of switching behaviour (i.e. the learning index, the HMM and the mixture model) 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 1**). The specific exclusion criteria for each analysis as well as the number of excluded sessions is described within the relevant section of the Methods and the results give the N for each analysis.

## ***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:

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

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 3E**.

## ***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:

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

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:

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

Where $1 \geq \pi_i \geq 0$ for all $\pi_i$, and $\Sigma_i$ πi = 1. Here, each $\beta_i$ reflects the survival parameter (average inter-switch interval) for each component distribution i and the $\pi_i$ 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 3E**. 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 sessions (of 607) were excluded, all in humans (11 total, including the 7 excluded previously because no switches were observed). 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:

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

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:

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

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

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

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).

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

Due to task differences, mice and humans had three possible states (two exploit states and one explore state), whereas monkeys had four possible states (three exploit states and one explore state) in Experiment 1. 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 three species, 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 an animal 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 transmission 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 Viterbi algorithm to discover the most probable posteriori 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 58 sessions out of 1056 (>5.5%, 27 for mice, 0 for monkeys, 13 for humans in **Experiment 1**, 8 for humans in **Experiment 2**, and 10 in humans in **Experiment 3**). 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 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 species, 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 $A_k$. A small number of sessions were excluded because their fitted HMM transition matrices did not admit a stationary distribution (**Experiment 1**: 49 out of 607 sessions; 29 for mice, 0 for monkeys, 20 for humans; **Experiment 2**: 7 of 150 sessions). For **Experiment 2**, this was in addition to the sessions excluded for not choosing all the available targets (6/150). We then took an average of these stationary distributions across all sessions for each species and plugged these back into the Boltzman equations to calculate the relative energy (depth) of exploration and exploitation as illustrated in **Figure 2F**.

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, $k_B$T 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 $k_B$T. 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 for all 3 species, see Methods section: Analysing HMM Dynamics (state dynamic landscapes) for more details. All statistical tests and statistical details were reported in the results.

## 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## Acknowledgments

The authors would like to thank Devin Heinz-Kehoe, Meriam 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

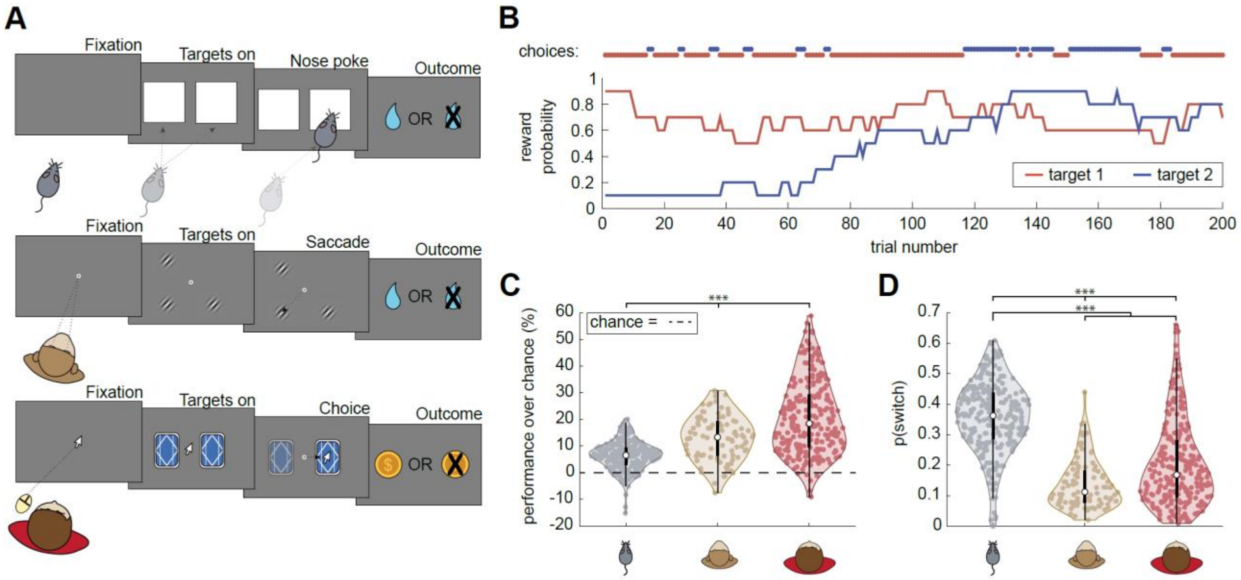

**Figure 1. Task design and behaviour across species**. **A**) A schematic representation of the bandit task in each species (mice = top, monkeys = middle, humans = bottom). **B**) Example reward schedule, including 200 trials from one session with one human. The reward probabilities of each of the 2 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. In figure, asterisks represent significance levels as follows: * indicates p < 0.05, ** indicates p < 0.001, and *** indicates p < 0.0001.

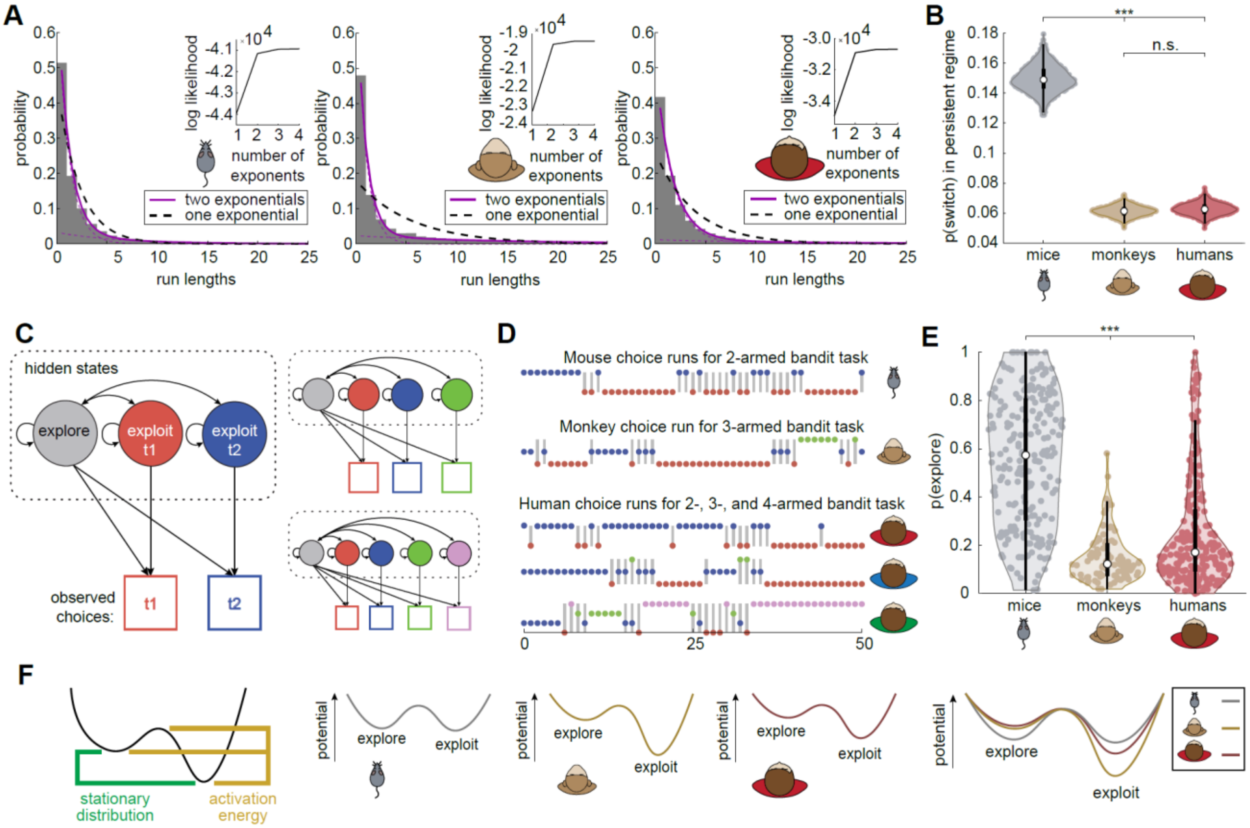

**Figure 2. Different patterns of switching and exploration across species. A**) Distributions of the number of trials between switch decisions ("run lengths") in mice, monkeys and humans. If the species had a fixed probability of switching, run lengths would be exponentially distributed (black dotted 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. In figure, asterisks represent significance levels as follows: * indicates p < 0.05, ** indicates p < 0.001, and *** indicates p < 0.0001.

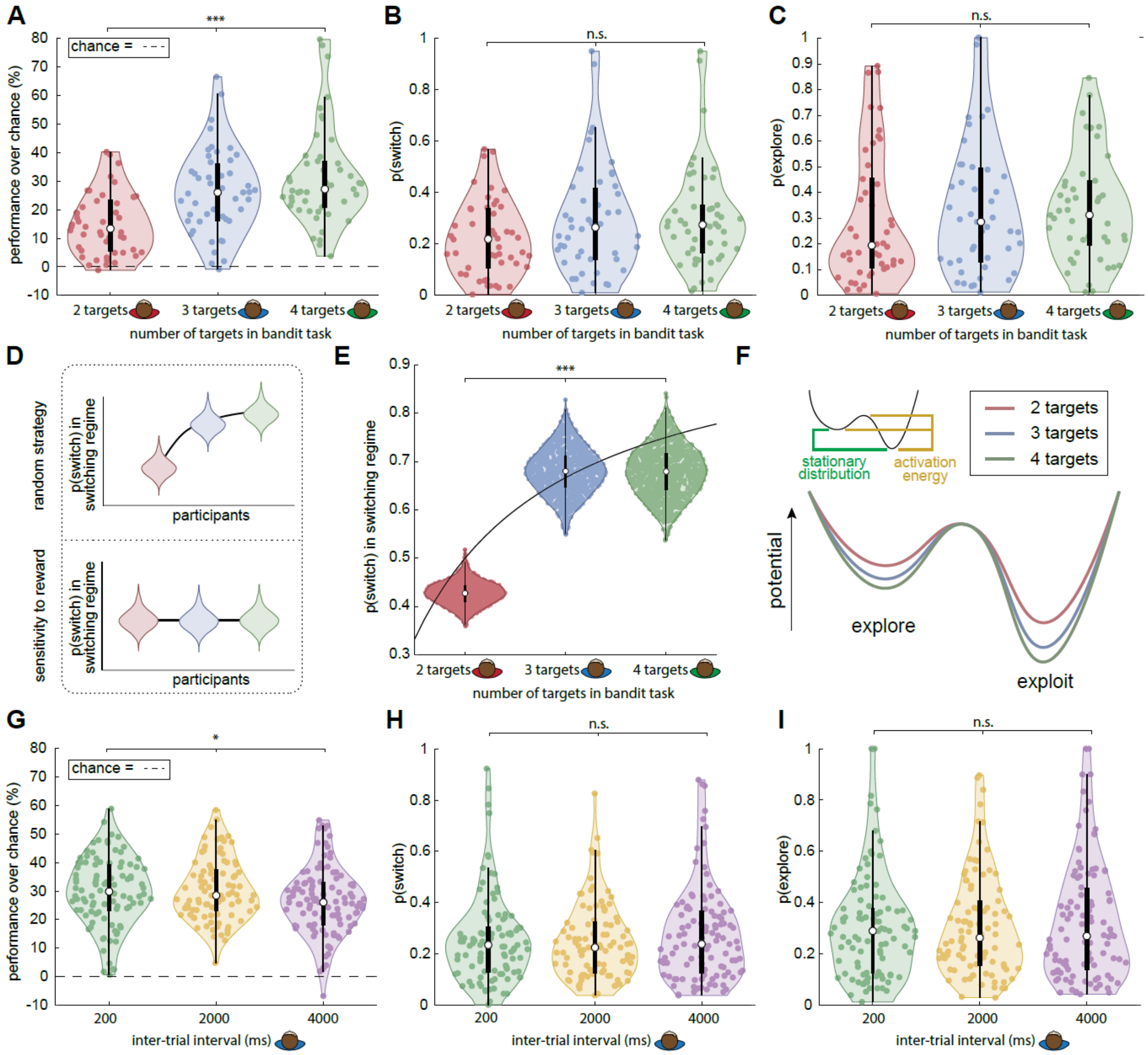

**Figure 3. Effects of manipulating the number of targets and the trial length in humans performing the bandit task (Experiment 2 and 3). A**) Percentage of reward relative to chance by number of targets (2, 3, or 4). Thick black lines = IQR, thin = whiskers, open circle = median. **B**) Switch probability by number of targets. **C**) Probability of exploration by number of targets. **D**) Cartoon illustrating predicted relationships between the switching-regime switch probability and the number of arms under the hypothesis of random exploration (top) or reward-dependent exploration (bottom). **E**) Switch probability for the fast-switching distribution (the "switching regime") by number of targets. **F**) State dynamic landscapes for varying numbers of targets (Same conventions as Figure 2F). **G-I**) Same as A-C across varying inter-trial interval times (200ms, 2000ms, 4000ms). Thick black lines = IQR, thin = whiskers, open circle = median. In figure, asterisks represent significance levels as follows: * indicates $p < 0.05$, ** indicates $p < 0.001$, and *** indicates $p < 0.0001$.

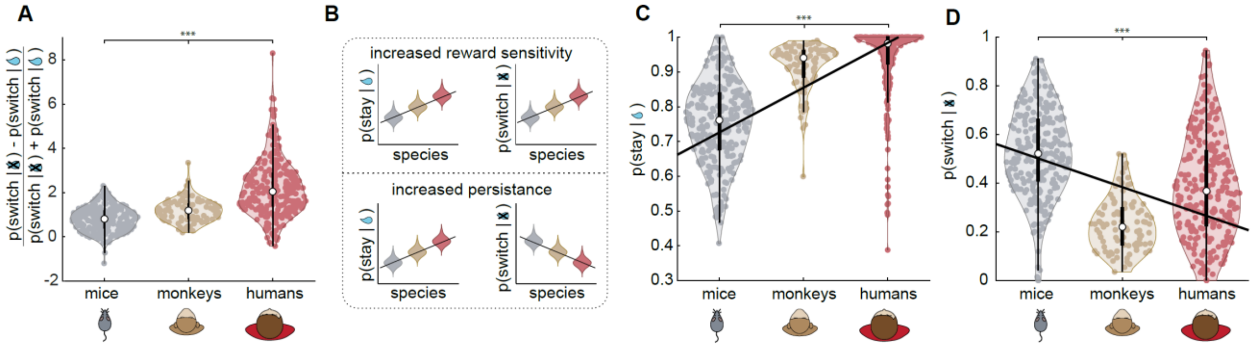

**Figure 4. Learning and persistence across species A**) Index of reward learning across species. Thick black lines = IQR, thin = whiskers, open circle = median. **B**) Hypothesis cartoon illustrating predictions under the hypothesis that species differences in switching were due to reward sensitivity (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. In figure, asterisks represent significance levels as follows: * indicates p < 0.05, ** indicates p < 0.001, and *** indicates p < 0.0001.