Divergent strategies for learning in males and females

Cathy S. Chen*[1], R. Becket Ebitz*[2], Sylvia R. Bindas[3], A. David Redish[2], Benjamin Y. Hayden[2], Nicola M. Grissom[+1]

*denotes equal contributions

1 Department of Psychology, University of Minnesota, Minneapolis MN 55455

2 Department of Neuroscience, University of Minnesota, Minneapolis MN 55455

3 Department of Behavioral Neuroscience, Oregon Health & Science University, Portland, OR 97239

+to whom correspondence should be addressed:

Nicola Grissom

Department of Psychology

University of Minnesota

75 East River Rd

Minneapolis, MN 55455

ngrissom@umn.edu

Abstract

Reported sex differences in decision-making and learning can be inconsistent across studies. One interpretation is that these sex differences are not driven by differences in ability, but by differences in strategy, which interact with task design. Here, we examined the strategies adopted by female and male mice as they learned the value of stimuli that varied across two dimensions. Female mice mastered image-value associations more quickly than male mice, and that they used a fundamentally different strategy to do so. Female mice constrained their decision-space early in learning. Conversely, male strategies changed frequently and were more influenced by the stochastic rewards. Individual strategies were related to sex-gated changes in neuronal activation in early learning. Together, we find that sex drives divergent strategies for learning, revealing substantial unrecognized variability in reward-guided decision-making and learning.

Introduction

Sex differences have been identified in many studies of reward-guided decision-making, seen across multiple mammalian species, and gender differences have been reported in humans (*1–5*). In some tasks males acquire more total rewards than females, a result that is sometimes interpreted as a male advantage in learning (*6*, *7*). However, seemingly minor changes to task paradigms can produce the opposite result. In fact, in some sequential probabilistic choice tasks, there is a female advantage in the number of total rewards earned (*2*, *8*, *9*). These findings suggest that males and females may not actually differ dramatically in their *ability to learn* from rewards. If so, then an alternative explanation is needed for the differences in sequential decision making that lead to different patterns of reward acquisition in different tasks (*10–12*).

Differences in reward outcomes need not solely arise from differences in ability to learn. Instead, they can also arise from differences in the choice of *what to learn about* (*10*, *12–15*): the strategy used for learning. In this view, sex-linked mechanisms influence learning strategies, and, because different strategies pay off differently in different environments, females have an advantage in some environments, while males have an advantage in others. However, this can only be uncovered in tasks that use choices that vary in multiple dimensions (*12*, *14*). To illustrate, one of the challenges of moving to a new city is finding a favorite restaurant. In a new dining scene, it is not always clear what dimensions of a restaurant--meaning features like location, price, or type of cuisine--best predict high quality meals. One strategy for learning about a new dining scene could be to try all restaurants at random, sampling the entire environment to simultaneously learn about all the dimensions in which restaurants vary, until a

winner is found. This strategy might find the best option, but would be incredibly time consuming and may be sensitive to noise in meal quality because each restaurant is sampled less frequently. Another strategy might be to try to learn about the feature dimensions by constraining the search in one dimension (like neighborhood) while learning about how restaurants vary in the other dimensions. To a naive observer, this approach may appear to be unnecessarily risk-averse or limited, but holding some dimensions constant can facilitate learning about other dimensions, particularly when feedback is noisy. However, it is not clear what factors influence how individuals select one of these different strategies.

To examine the possibility that sex differences in decision making may arise from different learning strategies, we examined male and female mice as they performed a two-dimensional decision-making task: a two-armed visual bandit (*11*, *16–22*). We considered the possibility that animals were adopting divergent strategies to solve the task, and that these strategies could be tuned by sex differences. We found that while males and females eventually reached the same performance level, female mice learned more rapidly than males and acquired more rewards over the course of learning. The difference in the rate of learning was not because females learned more from outcome of each trial, but because of a sex difference in the strategies that governed the next choice. Female mice adopted a consistent and systematic approach of preferring to choose options in one spatial location, which constrained the decision-space and accelerated their learning about image values. Conversely, choices made by males did not follow a single, straightforward strategic approach. Males appeared to consider information from both image and spatial dimensions simultaneously, and were highly sensitive to the stochastic experience of reward. As a result, individual males substantially changed *their own* choice strategies over

78 learning, differing from *themselves* much more so than did individual females. During early
79 learning, gene expression for the neuronal activation marker c-fos in the nucleus accumbens and
80 prefrontal cortex significantly correlated with the extent to which female animals (but not male)
81 used a systematic strategy. These results suggest that sequential decision making for reward can
82 be achieved through widely divergent strategies, within and between subjects, and that strategies
83 employed during learning are a significant source of sex differences in decision making.
84
85 **Results**
86
87 Age-matched male and female wildtype mice (n=32, 16 per sex, strain B6129SF1/J) were trained
88 to perform a visually-cued two-armed bandit task in touchscreen operant chambers (**Figure 1a**).
89 This task design was similar to those employed in humans and nonhuman primates (*11*, *16–24*),
90 in contrast to the spatial bandit designs frequently employed with rodents (*25–28*). Animals were
91 presented with a repeating set of two different image cues which were associated with different
92 probabilistic reward outcomes (**Figure 1b**).The reward schedule (80%/20%) was held constant
93 throughout the session. In contrast with spatial bandit designs, here reward contingencies were
94 yoked to image identity, which was randomized with respect to location in the chamber on each
95 trial. This means that the sides (left/right) where image cues appeared were not informative of
96 the reward contingencies. We repeated the task with six different sets of image pairs. Two out of
97 the six tested image pairs were excluded from the study due to extremely high initial preference
98 (>70%) for one image over another. We included four images pairs with equal initial preference
99 for each image and quantified behavioral data in bins of 150 trials for each animal.
100

Females showed accelerated reinforcement learning, but males and females reached equivalent final performance

To examine learning performance, we calculated in bins of 150 trials the average percentage of choosing the high-value image (23 bins in total). Regardless of sex, mice were capable of eventually learning which image was associated with the higher reward probability (**Figure 1c**, GLM, main effect of sex, $p = 0.51$, $\beta1 = -0.05$; main effect of number of trials, $p < 0.0001$, $\beta2 = 0.10$, see equation 1 in Methods). However, we repeatedly observed that females learned the image pair discrimination significantly faster than did males (GLM, interaction term, $p < 0.05$, $\beta3 = -0.02$). We compared these results to a deterministic version of the task in the same animals, in which one image was always rewarded (100%) and the other was never rewarded (0%). We did not find any significant sex difference in rate of learning across trials in the deterministic task (**Figure 1d**, GLM, interaction term, $p = 0.38$, $\beta1 = -0.004$, see equation 1 in Methods), suggesting the difference was revealed by the stochastic experience of reward.

To determine the origins of this sex difference in early performance, we first considered that sex differences in early learning might reflect differences in the rate of value-updating and/or the level of random noise in choice--the typical parameters of reinforcement learning models. We fit a delta-rule reinforcement model (*18*, *25*, *29*, *30*) to measure individual differences in the learning rate parameter and noise parameter, based on choices of images. However, the likelihood surface of the model given by parameters learning rate (α) and inverse temperature (β) was flat, which prevented parameter optimization. This suggested to us that the basic RL model based on images as the sole choice dimension could not characterize individual differences in learning in this task.

Since rodents are generally highly spatial, we hypothesized that mice might have a bias towards using spatial information earlier in the task, before switching to use image information, as demonstrated by the high rates of reward late in training. Consistent with our hypothesis, we observed a short period of high side bias in females early in learning (**Figure 1e**), which could include a preference for either the left or right side and seemed to precede the acquisition of the reward contingency. Following the decrease of side bias, female mice improved their percentage of choosing high-value image. (GLM, main effect of sex, $p < 0.001$, $\beta1 = -0.129$; main effect of number of trials, $p < 0.001$, $\beta2 = -0.017$, see equation 1 in methods). Strength of side bias independent of left or right side was calculated using methods described in previous behavioral lateralization literature (*31*).

Females systematically reduced the dimensions of the task by strongly preferring one side

A side bias is only one of several local strategies that mice could have been using as they learned the reward contingencies in the task. For example, another local strategy is a spatial win-stay strategy, where the side of the last choice is repeated if and only if it was rewarded. Alternatively, an image win-stay strategy would repeat the last image, if and only if it was rewarded, or an image bias strategy would simply select the previous image, regardless of reward. To understand how different animals employed these different local strategies and processed through them over time, we constructed a generalized linear model (GLM) to predict each choice, based on a weighted combination of these local strategies. The model had a term to account for two classes of basic strategies: outcome-independent strategies and outcome-dependent, win-stay, strategies (**Figure 2a**, see Equation 4 in methods). Outcome-independent

147 strategies (image repeat and spatial repeat) captured the tendency of repeating either the side or
148 the image of the previous trial, regardless of the outcome. Conversely, outcome-dependent
149 strategies followed a win-stay lose-shift policy for either a side or an image, which captured the
150 tendency to only repeat the side or the image of the previous trial when it was rewarded. Fitting
151 the GLM allowed us to estimate how much each of these four strategies was employed within
152 each animal on each bin of trials. We will call this set of weights--the precise pattern of local
153 strategies employed over time--the "global strategy" employed by each individual animal.
154
155 Across all animals, we found a global strategy where a specific procession through local
156 strategies was used when learning image pairs (**Figure 2b**). First, animals showed an early
157 tendency towards outcome-independent spatial repeat, giving way to a later interaction between
158 reward outcome and image choice, with outcome-insensitive image repeat (the optimal strategy)
159 increasing in the later stages of testing. To examine whether sex influenced the strength of this
160 strategy procession, we compared the global strategy used by male and female animals. We
161 observed a consistent and pronounced pattern of strategy procession only in females (**Figure 2c**).
162 In contrast, in males, the weight of both image-based strategies simply increased slowly over
163 time (**Figure 2d**), with little change in spatial strategies. Thus, neither a procession of multiple
164 strategies nor a prominent strategy in the early learning stage was observed in male mice.
165
166 The sex difference in this strategy selection was intriguing, but it could have been driven by only
167 a few females. Therefore, we next characterized individual variability in strategy via principal
168 component analysis. Here, we estimated the major axes of interindividual variability in strategy,
169 meaning in the unique combinations of the four strategy weight vectors over time and across all

170 animals, regardless of sex. Principal components (PC) 1 and 2 captured the majority of the
171 interindividual variance: 59% of the variability between animals (**Figure 2e**). PC1 reflected a
172 global preference for a side or an image and did not significantly differ between sexes ($p > 0.9$,
173 AUC = 0.43). The mean principal component scores of PC1 for females and males were 0.03 and
174 -0.03, respectively. The mean PC score difference between females and males ($mean_{(F-M)}$) was
175 0.07 (95% CI = [-1.70, 1.80], $t(30) = 0.08$). Critically, PC2 mirrored the same procession of
176 strategies observed in female, but not male mice (**Figure 2c-d**). This suggests that the extent to
177 which individuals used this procession of strategies explained a large fraction (22%) of the
178 interindividual variability in our animals. The mean PC score of PC2 for females was 0.98 and
179 for males was -0.98. The mean PC score difference between females and males ($mean_{(F-M)}$) was
180 1.96 (95% CI = [0.87, 3.05], $t(30) = 3.67$). Moreover, knowing the PC2 score (the projection of
181 an individual animal's behavior onto PC2) allowed us to discriminate male and female animals
182 with remarkable accuracy (receiver operating characteristic analysis, AUC = 0.86, significant
183 discrimination: $p < 0.001$). No other PCs differed between sexes ($p > 0.4$, AUC < 0.6). Together,
184 these results suggest that the choice of what strategy to follow explains substantial individual
185 variability in multidimensional decision-making, and that differences in strategy can depend on
186 sex.
187
188 There are two possible explanations why females consistently implemented the strategy
189 procession captured by PC2. One hypothesis is that this early side bias reflected an *energy saving*
190 strategy that saved time and/or effort by just repeating the same side. Alternatively, this early
191 side bias could be an *cognitively effortful* strategy to constrain decision-making to one
192 dimension. These two views make different predictions of the relationship between the use of

strategy procession captured by PC2 and reaction times (RTs), which were defined as the time between the onset of two visual stimuli and the registration of a nose poke response on one of the two stimuli. If side bias was an effort saving strategy, then the animals who score highest on PC2 should also make the fastest decisions. On the other hand, if side bias was a cognitively effortful strategy, the speed of decision-making should be slower in animals who use this strategy. To test these two hypotheses, correlation analyses were run to assess the relationship between the use of PC2 strategy and reaction time. PC2 scores were positively correlated with reaction time (**Figure 3a**, Spearman's correlation, $r_s = 0.452$, $p = 0.009$; Pearson's correlation, $r = 0.347$, $p = 0.051$), suggesting that the animals who used the early side bias strategy made slower decisions. This suggests that this strategy is effortful, rather than energy saving. There were no significant correlative relationships between reaction time and other PCs.

If the side bias in females was a cognitively effortful strategy, we would expect to see that female mice are slowest at making their choices when they are most engaging this strategy: during early learning. To test this hypothesis, we computed average RTs across 23 bins of 150 trials for males and females. Consistent with our hypothesis, females responded slower during early learning (bin 1-15) (GLM, interaction term, $\beta 3 = 0.03$, $p = 0.0007$) and significantly slower than males across all trials (GLM, main effect of sex, $\beta 1 = -0.62$, $p < 0.0001$). The mean RT across all trials for males was 1.89 seconds with standard deviation of 0.13, and the group average for females was 2.04 seconds with standard deviation of 0.21. The reaction time decreased as the animals ran more trials in both males and females (**Figure 3b**, GLM, main effect of number of trials, $\beta 2 = -0.04$, $p<0.0001$, see equation 1 in methods). Thus, female mice were slowest during the period in which they were using the side bias strategy the most, again

consistent with the idea that this is a cognitively effortful strategy, rather than an energy saving one.

Males varied strategies over time in response to immediate past reward

Although our analyses captured the procession of strategies employed across essentially all female mice that learned the task, they provided little insight into what the males were doing. One likely explanation is that males were more inconsistent than females. Males could be more inconsistent than females for any one of three reasons: hypothesis 1) males were more *random* (and thus each choice would be unpredictable within an animal), hypothesis 2) males were more *idiosyncratic* and less uniform as a group (and thus responses would differ across individuals, but still be predictable within an individual), or hypothesis 3) that males were more *changeable* (and thus a given male was predictable in the sense that he was not random, but he was still more likely to change his strategic approach from one epoch to the next).

To test hypothesis 1 (randomness), we asked whether male choices were more or less predictable than female choices. We reasoned that if males were just choosing randomly, it would be impossible to predict their choices. Therefore, within each block of trials in each animal, we calculated conditional mutual information (*32*, *33*), which quantifies the dependence of current choices (side, image) on the choice of previous trial, given the outcome of the previous trial. If the current choice an animal made was random, it would be independent of the choice and outcome of the previous trial, and we would expect to see low mutual information, shown as uniform "bands" on the probability heatmap (**Figure 4a**). Conversely, if the current choice was heavily influenced by the content of the previous trial, we would expect to see high mutual

information, shown in a more checkered or selective pattern on the probability heatmap. We calculated conditional mutual information for each trial bin across sexes. We found that mutual information decreased over time in both sexes, reflecting the gradual acquisition of the strategy of choosing high reward probability cue regardless of the outcome of the previous trial (**Figure 4a**, GLM, main effect of number of trials, $\beta 2 = -0.001$, $p = 0.0002$, see equation 1&5 in Methods). However, surprisingly, the mutual information of male mice was *higher* than that of females (main effect of sex, $\beta 1 = 0.043$, $p < 0.0001$), particularly early in learning (interaction term, $\beta 3 = -0.002$, $p < 0.0001$). Thus, males were, if anything, less random than females.

One possibility is that the high mutual information in males was a result of increased sensitivity to the last reward, meaning that the last reward had a bigger effect on the males' idiosyncratic decisions about what to do next. To estimate sensitivity to reward, we examined how reward outcomes affected the reaction time (RT) on the next trial in both males and females. If the decision of an animal was not affected by the outcome of the past trial, then we would expect to see no difference between reaction time for last rewarded and last unrewarded trial ($RT_{reward} - RT_{no\ reward} = 0$). Males responded significantly faster when they had just received a reward from the previous trial (**Figure 4b and 4c,** one-sample t-test, mean RT effect = -0.14, 95% CI = [-0.23, -0.05], $t(15) = -3.38$, $p = 0.004$). Conversely, the reaction times of females were not affected by the outcome of the last trial (one-sample t-test, mean RT effect = -0.03, 95% CI = [-0.13, 0.06], $t(15) = -0.75$, $p = 0.47$). These results again suggested that female decisions were not affected on a trial-by-trial basis by the outcome of each trial because they followed a global strategy while male choices that were heavily affected by recent rewards.

Although males were, if anything, less random in their decision-making than females, it remained possible that the apparent lack of local strategies occurred because male strategies were inconsistent--either because of idiosyncratic differences between males (hypothesis 2) or changeability within males (hypothesis 3). To do this, we developed a technique for comparing how similar one set of choices was to other set of choices. We expressed the choices in each bin as a probability vector, with each element of the vector reflecting the probability of that unique combination of behaviors {last choice, last outcome, this choice}. The average angle between any two of probability vectors across animals, trial bins, or image pairs is then a measure of the variability in choices between those two conditions. Males were not more idiosyncratic than females on a population level; that is, the choices of any male were not more variable from other males than any female's choices were from other females (**Figure 4d**, GLM, main effect of sex, $\beta1 = -1.47$, $p = 0.11$, see equation 1 in Methods). However, the males were more variable *within themselves,* both across bins within one image pair (**Figure 4e**, GLM, main effect of sex, $\beta1 = 4.24$, $p < 0.0001$, see equation 1 in Methods) and across multiple image pairs (**Figure 4f**, GLM, main effect of sex, $\beta1 = 4.54$, $p = 0.047$, see equation 1 in Methods). Overall, the variability in choices decreased across time, as the divergent strategies used by individual animals started to converge to the optimal strategy in this task, which is to choose the high-value image consistently (GLM, main effect of number of trials, within sex between suject: $\beta2 = -0.78$, $p < 0.0001$; within subject between bins: $\beta2 = -0.359$, $p < 0.0001$). Together, these results suggest that individual males displayed divergent choice patterns and were changing between complex strategies over time and the repetition of the same task, while females largely adopted a shared, systematic approach to learning.

To visualize animals' patterns of choices expressed in the probability vectors, we used multidimensional scaling (MDS) (*34–36*) to reduce the dimensionality of strategy space, allowing us to project the high-dimensional "strategy path" throughout learning onto a 2 dimensional space. This allows us to easily visualize the similarity between patterns of choice across animals over time and across repetitions. Each color path represents a strategy path used in a different repetition of the task (4 repetitions in total). **Figure 4g** shows examples of strategy paths of males and females. The optimal strategy in this bandit task, which is to choose the high value image consistently regardless of the outcome, is represented by a star in the low dimensional space. Both males and females "strategy path" showed gradual approximation to the optimal strategy over time. Consistent with the quantification described above, the strategy path of males are visibly more variable and different across repetitions of the task, whereas the strategy path of females were more unified and consistent across repetitions.

Sex mediated the ability of neuronal activity to explain strategy selection

The ability to learn and perform bandit tasks is highly sensitive to alterations in corticolimbic structures. However, it remains unclear how alterations in these structures predict choice strategy, much less sex differences in choice strategy. To address this question, we examined neuronal activity in several corticolimbic brain regions by examining the expression of *c-fos,* an immediate early gene often used as a marker of neuronal activation. The animals from the previous figures were sacrificed after the second day of a new, final image-reward pairing (each animal having completed 400-500 total trials), corresponding to when the female side bias was greatest. We compared mRNA expression level for *c-fos* across five brain regions, including nucleus accumbens (NAc), dorsal medial striatum (DMS), amygdala (AMY), hippocampus

(HPC), and prefrontal cortex (PFC), using quantitative real-time PCR (**Figure 5a**). In each of the five brain regions, females had a higher c-fos expression level than did males (unpaired t-test, NAc: $t(30) = 2.41$, $p = 0.02$; DMS: $t(30) = 2.31$, $p = 0.03$; AMY: $t(30) = 4.05$, $p < 0.001$; HPC: $t(30) = 2.74$, $p = 0.01$; PFC: $t(29) = 3.163$, $p = 0.003$).

To understand whether activation of any of these brain resions correlated with the side bias strategy, we constructed a GLM to predict PC2 from c-fos expression level in each brain region and sex. The results suggested that only two regions, the NAc and PFC predicted strategy use, as indexed by PC2 score (**Figure 5b**, GLM, NAc: $\beta1 = 0.72$, $p = 0.02$; DMS: $\beta2 = 0.48$, $p = 0.14$; AMY: $\beta3 = 0.52$, $p = 0.10$; HPC: $\beta4 = 0.55$, $p = 0.08$; PFC: $\beta5 = 0.75$, $p = 0.02$; sex was included as a variable in the model and was also significant: $\beta6 = 0.99$, $p = 0.0009$, see equation 2 in Methods). Because each region was also correlated with sex to differing extents (and sex independently predicted PC2), we next asked whether NAc and PFC were the best predictors of PC2 because these regions were the most strongly correlated with sex (**Figure 5c**). However, the predictive effect of NAc and PFC c-fos expression was not because NAc and PFC were the most highly with sex. Instead, sex was most strongly correlated with AMY, which was not a significant predictor of PC2. To confirm that these correlations between regional activation and early side bias strategy was meaningful, we fit the same GLM to predict PC1, and none of the predictor variables were significant. We confirmed these results with a Pearson product-moment correlation coefficient, which again suggested a significant positive correlation between c-fos expression in NAc/PFC and PC2 scores (**Figure 5d**, NAc: $r = 0.40$, $n = 32$, $p < 0.03$; PFC: $r = 0.41$, $n = 32$, $p < 0.02$; averages across a median split of PC2 within each sex are illustrated in **Figure 5e**; main effect sex: NAc: $F(1,28) = 12.87$, $p = 0.001$; PFC: $F(1,28) = 13.47$, $p = 0.001$).

Next, we asked whether an animals' sex altered the relationship between NAc and PFC c-fos activity PC2 scores. To do this, we used a structural equation modeling (SEM) approach (*37*, *38*) to analyze the structural relationship between sex, gene expression, and PC2 and latent constructs (**Figure 5f**). First we used a direct model and regressed c-fos expression of either NAc or PFC on strategy selection, both NAc and PFC were significant direct predictors of PC2 scores (NAc: $\beta = 0.72$, $p = 0.022$; PFC: $\beta = 0.75$, $p = 0.019$, see equation 6 in Methods). Then we fit a mediation model that allows us to understand how sex influences neural activation in NAc and PFC, which in turn influences strategy selection. Regressing the mediator variable sex on c-fos expression in NAc/PFC confirmed that neural activation is a significant predicor of the mediator sex (NAc: $\alpha = 0.20$, $p = 0.024$; PFC: $\alpha = 0.26$, $p < 0.004$, see equation 7 in Methods). When we regressed strategy selection on both the mediator variable (sex) and independent variable (neural activation in NAc/PFC), the result showed that the mediator sex was a significant predictor of strategy selection (NAc: $\beta 1 = 1.66$, $p = 0.008$; PFC: $\beta 1 = 1.64$, $p = 0.012$), and the strength of the direct model is now greatly reduced and became non-significant when accounted for the mediating effect of sex (NAc: $\beta' = 0.38$, $p = 0.20$; PFC: $\beta' = 0.33$, $p = 0.31$). The Sobel (1982) first-order test was used to assess the presence of mediation (*38*). The indirect effect was calculated as the product of coefficients and was significant for both NAc and PFC (NAc: $\alpha\beta' = 0.34$, $z = 1.836$, $p < 0.039$; PFC: $\alpha\beta' = 0.42$, $z = 2.035$, $p < 0.026$). Together, these results suggest that the relationship between PFC and NAc c-fos and PC2 differed, depending on the animals' sex. This suggests that sex-linked mechanisms gate the relationship between these circuits and strategic decision-making and highlight these regions as promising targets for future

353 studies looking at the effects of sex on the neural circuits responsible for implementing strategic
354 learning.
355
356 **Discussion**
357
358 By training male and female mice on a stochastic two-dimensional decision-making task, we
359 were able to evoke a range of problem solving strategies across individuals. In this task, each cue
360 has two dimensions - the identity of the image and the location of the image. Animals had to
361 explore the reward value associated with both cue dimensions to determine which were most
362 predictive of reward. Although both male and female mice eventually learned the right strategy,
363 choosing the high-value image, female mice learned faster. The richness of this task allowed us
364 to uncover sex differences in *how* the animals achieved the associations across time. We
365 discovered that female mice adopted a consistent and systematic approach where they processed
366 through different strategies over time. Early in learning, they constrained their search space by
367 only sampling the outcomes of images on one side (left or right). This approach, which occurred
368 when animals were most uncertain about the best choice, reduced the number of dimensions they
369 were learning about and permitted more rapid acquisition of the image-value association. In
370 contrast, males employed a strategy of decisions that seemed to combine both image and spatial
371 location, changed frequently, and was strongly influenced by the immediate prior experience of
372 reinforcement. While both sexes eventually reached equivalent levels of performance, our data
373 reveal that the journeys individual animals take to get there can vary dramatically, implicating
374 the potential for wide divergence in neural circuit mechanisms in normal decision making.
375

One fundamental unanswered question is *why* females as a group employed a highly similar and consistent strategy. Zador (2019) recently proposed that much of animal behavior is not dictated by supervised or unsupervised learning algorithms, but are largely innate, shaped by biological constraints (*39*). The biological constraints and organization imposed by the multiple mechanisms of sexual differentiation are known to drive a tuning of the circuits important for reward-guided decisions (*40–44*). Among these are well-known effects of testosterone in driving exploratory and impulsive behavior (*43*, *45*, *46*). Conversely, energy-conserving and habitual behaviors are more prevalent in female animals, including during foraging (*1*, *2*, *47*). Gonadal hormones, such as ovarian hormone estradiol (E2), are thought to exert modulatory control over cost/benefit decision-making that increased E2 resulted in reduction of high-effort choices (*5*, *48*, *49*). In addition, dissociable impacts of sex chromosomes on reward-guided behaviors (*50*) that have been described as promoting habit in XX carriers and increased effort in XY carriers. Of course, most impacts of sexual differentiation are graded, rather than dichotomous across the sexes. Indeed, here we found that a small number of males showed some tendency to use the female strategy, implicating graded mechanisms of masculinization in sex-gated strategy selection. Previous evidence of sex-differences in decision-making has been interpreted as evidence that females avoid unnecessary effort in the pursuit of food. However, our results suggest that this strategy may be effortful, not effortless. Further, in many circumstances, it may be a better strategy than the male pattern of indiscriminate exploration. Thus, it is possible that these effects are due to differences in the behavioral ecology of male and female animals, which creates different biological constraints on learning across sexes.

What kinds of environments might advantage the typical female strategy over the typical male strategy or vice versa? Our result suggests that the systematic approach employed by female mice may have greater success with high-dimensional learning tasks. By reducing or eliminating choices in one dimension, females decreased the number of dimensions that varied at a time, and were able to learn other associations more quickly. However, in sparser environment, a strategy of exhausting all options in one dimension could become less effective. Having a range of divergent strategies and changing choice patterns in response to the reward outcome, as seen in male mice, may lead to a greater chance of success in varying and volatile environments (*17*, *51*, *52*) at the cost of greater risk to an individual male. Indeed, it is possible that these differences in the match between sex-specific strategies and the environment may be a major contributing factor in the inconsistent gender and sex differences across tasks with different levels of volatility (*1*, *2*). An intriguing possibility is that the unified, consistent, and systematic strategy we observed in female mice, as well as the volatile and diverse strategy we observed in male mice in the same task, may emerge from evolved sex-biased strategies for foraging in the wild that were critical to survival for the species as a whole, by dividing risk and reward across the population.

Sequential decision-making and learning is often studied with spatial bandit tasks, in which reward probabilities are linked to left and right levers or sides that are visually identical (*17*, *18*, *25*, *53–55*), particularly in rodent models. In these spatial bandit tasks, side bias in choice has sometimes been reported in rodent operant work as a behavioral artifact and animals displaying such bias were often excluded from experiments (*56–58*). However, in the current task, both the side and the identity of the image cues could have been informative of the reward probabilities.

421 In principle, animals could simultaneously sample both dimensions to learn side values and
422 image values at the same time. However, in practice, it appears that the early side bias in female
423 mice "jump-started" their learning by controlling for space while exploring choice-outcome
424 values of the images, which in this task happened to be the more informative dimension of
425 reward. Intriguingly, this suggests that females were covertly learning about the correct cue
426 dimension while behaviorally selecting the wrong item, and were able to convert this to
427 successful learning due to the stability of the task structure. This view suggests that , we should
428 be able to design tasks that prevent the successful use of this strategy, and which might therefore
429 shift the presentation of the sex difference in decision-making.
430
431 Our data implicate the prefrontal cortex (PFC) and nucleus accumbens (NAc), part of the ventral
432 striatum, in the differences in strategy between males and females. These regions have been
433 widely implicated in reward-guided decision making, but so have the other regions we tested for
434 which we didn't find a significant relationship to behavior (*11*, *12*, *15*). One possibility is that the
435 PFC and accumbens are particularly engaged in strategic decision-making. This resonates with
436 previous studies that have implicated the PFC in implementing strategies and rule-guided
437 behaviors (*51*, *55*, *59–63*) and the NAc in selecting and implementing learning strategies (*12*,
438 *14*). Implementing different strategies produces changes in how different choice dimensions are
439 represented in the PFC and NAc (*64*), and lesions in the NAc can drive animals towards a low-
440 dimensional action-based strategy or prevent animals from switching between strategies (*11*, *14*).
441 The PFC is also sensitive to gonadal hormones during risky decision making (*65*), and
442 dopaminergic function in the accumbens regulates risky decision making in a sex-specific
443 manner (*66*), perhaps due to sex differences in dopamine neurons (*44*). Here, the relationship

444 between both PFC and NAc and strategy use was mediated by sex, suggesting that whatever the

445 relationship between these regions and strategic decision-making, it is likely to be sex-specific.

Methods

Animals. 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 while being mildly food restricted (85-95% of free feeding weight) for the experiment. Animals engaging in operant testing were housed in a 0900–2100 hours reversed light cycle to permit testing during the dark period, between 09:00 am and 5:00 pm. Before operant chamber training, animals were food restricted to 85%-90% of free feeding body weight and had been pre-exposed to the reinforcer (Ensure). Pre-exposure to the reinforcer occurred by providing an additional water bottle containing Ensure for 24 hours in the home cage and verifying consumption by all cagemates. Operant testing occurred five days per week (Monday-Friday), and the animals were fed after training with ad lib food access provided on Fridays. All animals were cared for according to the guidelines of the National Institution of Health and the University of Minnesota.

Apparatus. Sixteen identical triangular touchscreen operant chambers (Lafayette Instrument Co., Lafayette, IN) were used for training and testing. Two walls black were acrylic plastic. The third wall housed the touchscreen and was positioned directly opposite the magazine. The magazine provided liquid reinforcer (Ensure) delivered by a peristaltic pump, typically 7ul (280 ms pump duration). ABET-II software (Lafayette Instrument Co., Lafayette, IN) was used to program operant schedules and to analyze all data from training and testing.

Operant Training

468 *Pretraining.* animals were exposed daily to a 30-min session of initial touch training, during
469 which a blank white square (cue) was presented on one side of the touchscreen, counterbalancing
470 left and right between trials. This schedule provided free reinforcement every 30 seconds, during
471 which the cue was on. If animals touched the cue during this period, a reward three times the
472 size of the regular reward was dispensed (840 ms). This led to rapid acquisition. Following this,
473 animals were exposed daily to a 30-min session of must touch training. This schedule followed
474 the same procedure as the initial touch training, but free reinforcers were terminated and animals
475 were required to nose poke the image in order to obtain a regular reward (7-uL, 280 ms).
476
477 *Deterministic pairwise discrimination training.* Animals were exposed to 10 days of pairwise
478 discrimination training, during which animals were presented with two highly discriminable
479 image cues ("marbles" and "fan"). One image was always rewarded and the other one was not.
480 Within each session, animals completed either 250 trials or spent a maximum of two hours in the
481 operant chamber (typically these mice completed ~200 trials/day).
482
483 *Two-armed bandit task.* Animals were trained to perform a two-arm visual bandit task in the
484 touchscreen operant chamber. Each trial, animals were presented with a repeating set of two
485 different images on the left and right side of screen, counterbalancing left and right across the
486 session. Nose poke to one of the displayed images on the touchscreen was required to register a
487 response. Nose poke on one image triggered a reward 80% of the time (high payoff image),
488 whereas the other image was only reinforced 20% of the time (low payoff image). Following the
489 reward collection, which was registered as entry and exit of the feeder hole, the magazine would
490 illuminate again and the mouse must re-enter and exit the feeder hole to initiate the next image

trial. If the previous trial was unrewarded, a 3-second time-out was triggered, during which no action could be taken. Following the timeout, the magazine would illuminate and the mouse must enter and exit the feeder hole to initiate the next image trial. The ABET II system recorded trial to trial image chosen history, reward history, grid position of the images with time-stamp. Within each day, animals completed either 250 trials or spent a maximum of two hours in the operant chamber. Animals were given 14 days to learn about the probabilistic reward schedule of one image pair, before moving onto the next image pair. A total of six image pairs were trained, but two image pairs were eliminated from analyses due to very high initial preference (>70%) for one novel image over another, indicating that (to the mice) these images appeared unexpectedly similar to previously experienced images with learned reward values.

RNA quantification. At the end of training, animals were sacrificed after the second day of learning a new image pair (around 400-500 trials of experience per mouse), when we expected to see the biggest difference in learning performance and strength of lateralization. Animal brains were extracted and targeted brain regions were dissected. We extracted RNA from targeted brain areas and assessed gene expression for the *fos* genes in the nucleus accumbens (NAc), dorsal medial striatum (DMS), amygdala (AMY), and hippocampus (HPC), using quantitative Real Time PCR system (BioRad, USA). Fos expression normalized to the housekeeper gene glyceraldehyde 3-phosphate dehydrogenase (*gapdh*) was calculated using the comparative delta Ct method.

Data analysis

Generalized Linear Models (GLMs). In order to determine whether sex and number of trials (bins) predicts the accuracy of the task, strength of lateralization, reaction time, mutual information (MI), or angle between probability vectors, we fit a series of generalized linear models of the following form:

$$Y = \beta_0 + \beta_1(sex) + \beta_2(trials) + \beta_3(sex)(trials) \quad [1]$$

Where Y is the dependent variable (accuracy, laterality, reaction time, MI, or angle). In this model, β1 described the main effect of sex and β2 described the main effect of number of trials (bins). β3 captures any interaction effect between sex and number of trials (bins).

To determine whether c-fos expression in NAc, DMS, AMY, HPC, PFC, and sex predicted the weights of Principal Component (PC) 2, we fit the following generalized linear model.

$$PC2 = \beta_0 + \beta_1(NAc) + \beta_2(DMS) + \beta_3(AMY) + \beta_4(HPC) + \beta_5(PFC) + \beta_6(sex) \quad [2]$$

In this model, β1- β5 captures the predictive effect of gene expression in five regions on the use of PC2 strategy. β6 described the effect of sex on the weights of PC2.

Degree of lateralization. As a measure of the strength of side bias, we used the absolute percentage of laterality (*31*), calculated for each mouse according to the following formula:

$$Degree\,of\,laterality = \left|\frac{right - left}{right + left}\right| \quad [3]$$

535 *Generalized Logistic Regression Model.* Mice could base their decisions on reward history in the
536 spatial or image domains or on choice history in the spatial or image domains. To determine how
537 these four aspects of previous experience affected choice and how these effects changed over
538 time, we estimated the effect of the last trials' reward outcome (O), image choice (I), and chosen
539 side (S) using logistic regression. If image (image 1) was on the left side of the screen, we could
540 predict the probability of choosing that image as a linear combination of the following four
541 terms:

$$\begin{aligned} log(\frac{p(I_{1,t})}{p(I_{2,t})}) =& \beta_0 + \beta_1 * (I_{1,t-1} - I_{2,t-1}) + \\ & \beta_2 * O_{t-1} * (I_{1,t-1} - I_{2,t-1}) + \\ & \beta_3 * (S_{L,t-1} - S_{R,t-1}) + \\ & \beta_4 * O_{t-1} * (S_{R,t-1} - S_{L,t-1}) \end{aligned} \quad [4]$$

542

543 Where each term (O, I, and S) is a logical, indicating whether or not that event occurred on the
544 last trial. As a result, the term (I_{1, t-1} - I_{2, t-1}) is 1 if image 1 was chosen on the last trial,
545 but -1 otherwise. The term β1 thus captures the tendency to either repeat the previous image
546 (when positive) or choose the other image (when negative). The term β2 O_{t-1} accounts for
547 any additional effect of the previous image on choice, when that previous choice was rewarded..
548 If image 1 was on the left side, (S_{L, t-1} denotes the probability of repeating the left side
549 where image 1 appeared. However, because image 1 could be either on the left or the right side
550 of the screen (which allowed us to dissociably estimate the probability of choosing it based on
551 side bias or image bias), we expanded the (S_{L, t-1} - S_{R, t-1}) term to account for the
552 current position of image 1 as follows:

553

554 $$...((I_{1,t} = L)(S_{L,t-1} - S_{R,t-1}) + (I_{1,t} = R)(S_{R,t-1} - S_{L,t-1}))...$$

555 Meaning that the current position of image 1 determined the sign of the side bias term. This
556 model was fit individually to each bin of 150 trials, within each animal and image pair, via cross-
557 entropy minimization with a regularization term (L2/ridge regression).
558
559 *Principal component analysis.* In order to determine how decision-making strategies differed
560 across animals and bins, we looked for the major dimensions of interindividual variability in
561 decision-making strategies. To do this, we took advantage of the fact that the coefficients of the
562 generalized linear model provided a simplified description of how decision-making depended on
563 image, side, and outcome for each subject in each bin. Because the generalized logistic
564 regression model estimated 4 terms per image pair and there were 23 independent bins per image
565 pair, this meant that each animals' behavior for a given image pair could be described as a 4*23
566 by 1 dimensional vector. We then used principal component analysis to identify the linear
567 combinations of model parameters that explained the most variance across subjects and
568 repetitions of image pairs (across 32 (animals) x 4 (repetitions) = 128 total strategy vectors). The
569 first two principal components, which explained the majority of the variance (59%), are
570 illustrated in Figure 2e.
571
572 *Conditional mutual information and model-free analyses.* To account for idiosyncratic strategies,
573 which could vary across animals or image pairs, we used a model-free approach to quantify the
574 extent to which behavior was structured without making strong assumptions about what form
575 this structure might take. We quantified the extent to which choice history was informative about
576 current choices as the conditional mutual information between the current choice (C) and the last
577 choice (C_t-1), conditioned on the reward outcome of the last trial (R):

$$I(C_t; C_{t-1}|R) = \sum_{r \in R} \sum_{c_{t-1} \in C} \sum_{c_t \in C} P_{C_t, C_{t-1}, R}(c_t, c_{t-1}, r) log \frac{P_R(r) P_{C_t, C_{t-1}, R}(c_t, c_{t-1}, r)}{P_{C_t, R}(c_t, r) P_{C_{t-1}, R}(c_{t-1}, r)} \quad [5]$$

Where the set of choice options (C) represented the unique combinations of each of the 2 images and 2 sides (4 combinations). To account for observed differences in overall probability of reward for male and female animals, the mutual information was calculated independently for trials following reward delivery and omission, and then summed across these two conditions.

We used a similar approach to provide a model-free description of the animals' choice patterns. Briefly, instead of finding the set of beta weights that best described reliance on various history-dependent strategies over time, we directly calculated the joint probability of each possibility combination of last choice (image and side), last outcome (reward and unrewarded), and current choice (image and side). This means that we represented the animals' history-dependent choice pattern for each image pair as an 32-dimensional vector (4 (last choice) x 2 (last outcome) x 4 (current choice) = 32) of joint probabilities. Via a geometric interpretation of a multinomial distribution, we considered the animal's pattern of behavior within any bin of trials as a point on the 32-1 dimensional simplex formed by length-1 vectors. This geometric approach allowed us to map strategies over time or across bins as a diffusion process across this simplex, where the angle between two vectors (between animals/between bins/between repetitions) is proportional to step between them on a strategy simplex. The bigger the step between two vectors, the more variable the behavior pattern is.

599 *Mediation Analysis.* First we used a direct model and regressed c-fos expression of either NAc or
600 PFC on weights of PC2. When assessing a mediation effect, three regression models are
601 examined:
602
603 Model 1 (direct):

$$PC2 = \gamma_1 + \beta(NAc) + \epsilon 1 \quad [6]$$

605 Model 2 (mediation):

$$Sex = \gamma_2 + \alpha(NAc) + \epsilon 2 \quad [7]$$

607 Model 3 (indirect)

$$PC2 = \gamma_3 + \beta'(NAc) + \beta_1(sex) + \epsilon 3 \quad [8]$$

609
610 In these models, γ1, γ2, and γ3 represent the intercepts for each model, while ε1, ε2, and ε3
611 represent the error term. β denotes the relationship between dependent variable (PC2 weights)
612 and independent variable (NAc c-fos expression) in the first model, and β'denotes the same
613 relationship in the third model. α represents the relationship between independent variable (NAc
614 c-fos expression) and mediator (sex) in the second model. The mediation effect is calculated
615 using the product of coefficients (αβ1). The Sobel test is used to determine whether the
616 mediation effect is statistically significant (*38*).
617
618 *Reinforcement Learning Model.*
619 We considered a basic reinforcement learning model, following the Rescorla Wagner rule. In this
620 model, subjects first learn the expected value of each image based on the history of its previous

621 outcome value Q and use these Q values to decide what to do next. The expected value of arm *k*

622 on the *t* th trial, Q_t^k, is updated based on the reward outcome of each trial:

623 $$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$$

624 In each trial, $r_t - Q_t^k$ captures the reward prediction error (RPE), which is the diffe-rence

625 between expected outcome and the actual outcome. The parameter α is the learning rate, which

626 determines the rate of updating RPE. Action selection was performed based on a Softmax

627 probability distribution:

628 $$P(a_{t+1} = k) = \frac{e^{\beta Q_t^k}}{\sum_j e^{\beta Q_t^j}}$$

629 where inverse temperature β determines the level of random exploration.

630

631

632

633

634

635

636

637

638

639 **Acknowledgements**

640

641 The authors would like to thank Sarah Heilbronner and Vincent Costa for helpful comments on
642 the manuscript. This work was supported by startup funds from the University of Minnesota
643 (NMG), a NIMH T32 training grant (MH115886), and a NARSAD Young Investigator Grant
644 (RBE).

Figures and Figure Captions

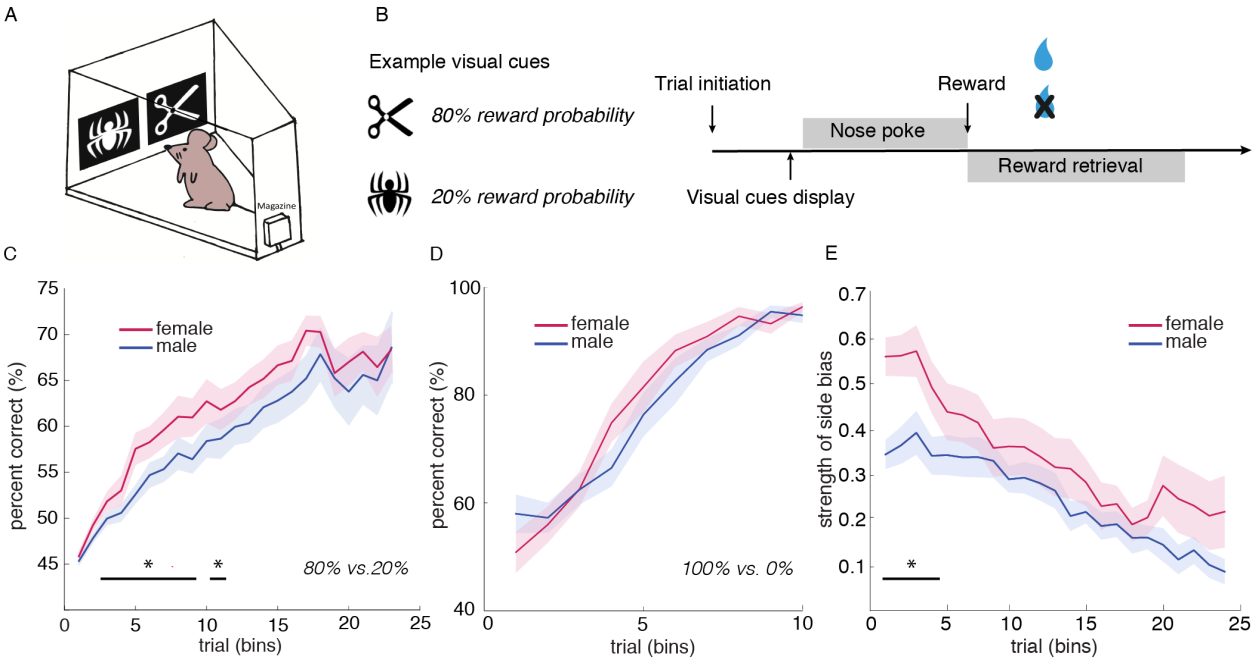
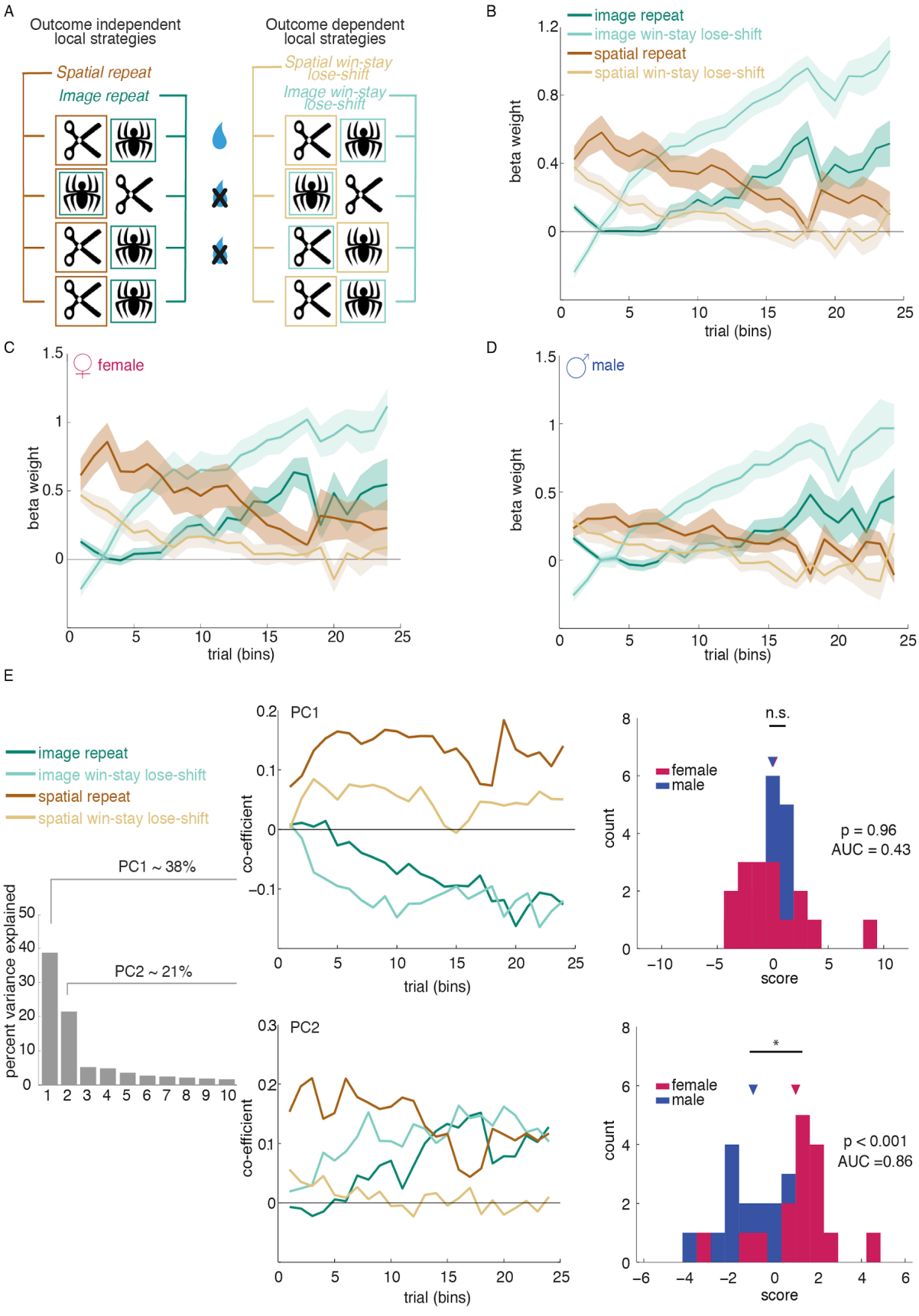
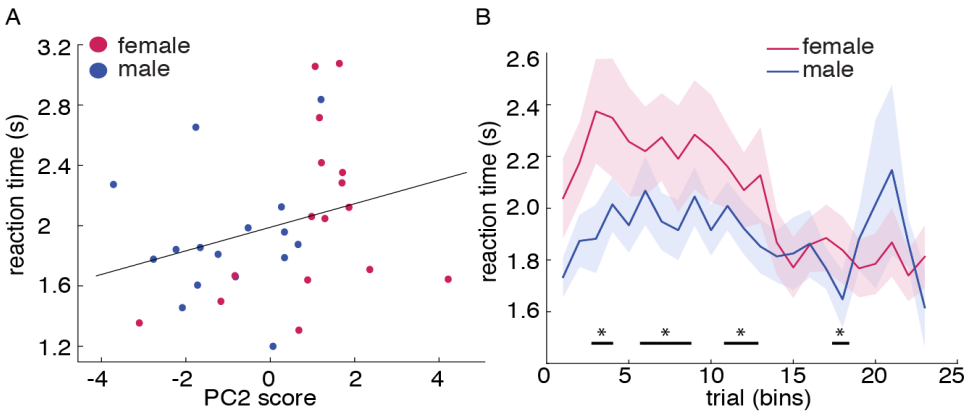


Figure 1. **Females showed accelerated acquisition of the high reward probability image in a stochastic two-armed visual bandit task.** A) Schematic of the mouse touch-screen operant chamber used in our task. B) Schematic of two-armed Visual Bandit task. Images varied between the two locations across trials. The reward probabilities for two images are 80% and 20%, respectively. C) Average learning performance (percent correct) across four repetitions of the task with four pairs of images. While both males and females reached the same final performance, females displayed an accelerated learning curve. D) No sex difference in learning performance was observed in deterministic reward schedule (100%/0%). E) Females displayed stronger side bias for item selection on the touchscreen, regardless of the direction of lateralization, early on in learning. This behavior lateralization disappeared as female mice learned the task. Data shown as bins of 150 trials. * indicates $p < 0.05$. Bars ± SEM. N=16/sex, wildtype F1 strain B6129SF1/J.



660

Figure 2. Female mice use a procession of strategies, initially using a spatial bias followed by a switch to responding based on image domain. A) Schematic of four basic local strategies based on choice and reward history of image and spatial dimensions of the task. B) A generalized logistic regression model revealed a global strategy - a clear procession of four local strategies - that mice started repeating one side before switching to choosing the reinforced image. C) female mice displayed more pronounced global strategy procession from spatial-based strategy to image outcome-based strategy. D) male mice displayed increased image win-stay lose-shift over time but no prominent global strategy in the early learning stage was observed. E) A principal component analysis (PCA) was conducted on the estimates of global strategy strength over time across all animals regardless of sex. Principal component (PC) 1 and 2 accounted for about 60% of the variance. PC 1 described a general preference for responding based on image value and did not differ between sexes. PC 2 captured the same global strategy procession reflected in the generalized logistic model - a strong contribution of the "side repeat" behavior early in training, followed by a rapid transition to "image outcome", indicative of a sudden shift away from "where" and towards "what" in solving the task. Projecting each animal onto this PC 2 showed a clear separation between the sexes (blue male, pink female), AUC = 0.86, $p < 0.001$. This suggests that the strategy procession from spatial repeat to image outcome is a female-specific strategy. Note that a few males are positive for Principal Component 2, and their individual behavior supports that these males also employed this strategy to a weaker extent. In contrast, the few females that are negative for this Principal Component did not show evidence of having learned the task. Data shown as bins of 150 trials. Bars ± SEM.



682

683 **Figure 3. The side-to-image strategy procession captured by Principal Component (PC) 2 is**

684 **a cognitively demanding strategy but not a time-saving strategy.** A) Correlation analyses

685 revealed a significant positive correlation between PC2 scores and reaction time. The decision-

686 making time was longer within animals primarily used PC2 strategy. B) Predominantly using

687 PC2, females responded slower during early learning (GLM, interaction term, $\beta 3 = 0.03$, $p =$

688 0.0007). Note that, the slow reaction time during early learning in females matches with the time

689 period (bin 1-10) during which females relied on side-bias "heuristics" for decision-making.

690 Data shown as bins of 150 trials. * indicates $p < 0.05$. Bars ± SEM.

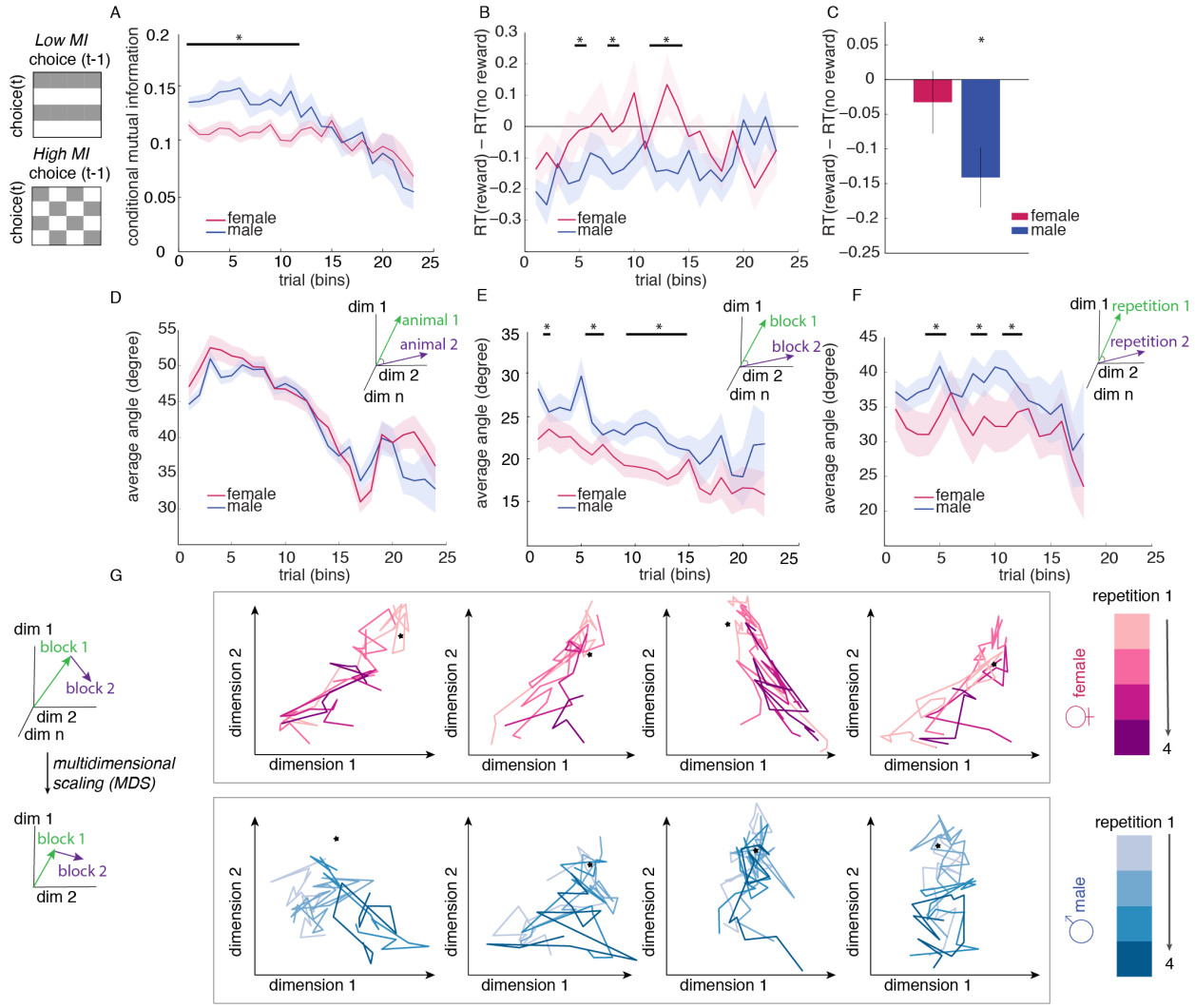
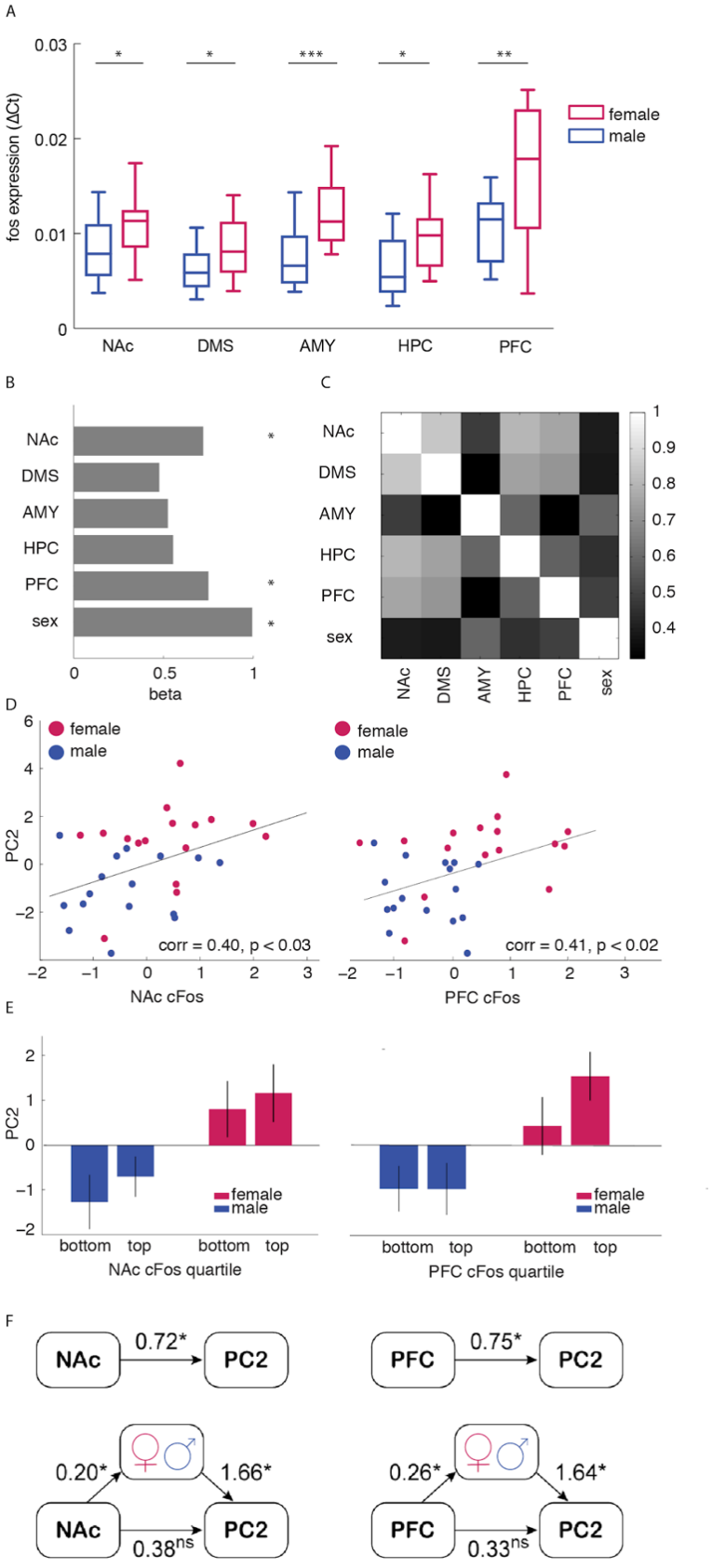


Figure 4. Male mice were more likely to differ from themselves over time, with choice patterns dependent on past outcomes. We expressed each animal's history-dependent choice pattern as an 32-dimensional vector of joint probabilities and measured the angle between vectors, which is proportional to the step between them on a strategy simplex. A) Illustration of choice patterns of low mutual information and high mutual information. If choice on trial *t* is independent of choice on the previous trial (*t-1*), probability heatmap should show band-like pattern (choosing the same choice regardless of the previous choice). Conversely, high mutual information has more checkered choice patterns. Conditioned mutual information is higher in males, indicating that responses are more uniquely affected by the previous trial variables than

they are in females. B) One-sample t-test was conducted across bins to compare the difference in reaction time (RT) between rewarded and unrewarded trials to 0 (when there is no effect of past outcome on the reaction time). Male mice have significant RT effects on the last reward. There was no difference in reaction time between rewarded and unrewarded trials in female mice. The bins marked by asterisks have $p < 0.05$ for the one-sample t-test. C) average RT effect of last reward across all trials. Overall, male responded faster when the last trial was rewarded than unrewarded. D). Males and females were equally variable between animals within sex. An individual male is no more different from other males in behavior than a female is from other females. E) Choice patterns of a given male compared to himself over bins of 150 trials were more variable than in a given female compared to herself . F) Choice patterns of a given male to himself were more variable and divergent across repetitions of the same task than in females compared to themselves across repetitions. G) Multidimensional scaling (MDS) was used to visualize animal's strategy path across trials and repetitions by reducing the dimensionality of the strategy space. Each of the four colors within one sex represents one repetition of the task. The star represents the point of optimal strategy for this task, which is to choose the high reward probability image. In both males and females, the strategy paths showed a gradual approach to the optimal strategy point, indicating that both sexes were able to learn the optimal strategy. The strategy paths of females are consistent and similar across repetitions, suggesting that female mice used a similar strategy every time to solve the problem. On the other hand, in male mice, the steps between each bins of 150 trials in the strategy space were larger, suggesting higher variability in choice patterns. Thus, male mice used divergent strategies throughout learning and used different approaches each time to learn the same task. Data shown as bins of 150 trials. * indicates $p < 0.05$. Bars ± SEM.



724

Figure 5. Both sex and neuronal activity can account for strategy selection, but sex mediated the ability of neural activity to explain strategy selection.

A) cFos gene expression (qRT-PCR) in five brain regions: nucleus accumbens (NAc), dorsal medial striatum (DMS), amygdala (AMY), hippocampus (HPC), and prefrontal cortex (PFC). Female mice showed elevated c-fos expression across all five brain regions. Asterisks marked significant difference between sexes (*: $p < 0.05$ **: $p < 0.01$ ***: $p<0.001$). **B**) Heatmap of correlation matrix of c-fos expression level among five brain regions. **C**) c-fos expression in NAc and PFC, and sex, predict the use of PC2 strategy (GLM, NAc: $\beta1 = 0.72$, $p = 0.02$; PFC: $\beta5 = 0.75$, $p = 0.02$; sex: $\beta6 = 0.99$, $p = 0.0009$). Asterisks marked significant beta weights ($p < 0.05$). **D**) cFos gene expression in NAc and PFC is significantly correlated with the weight of PC2. **E**) The use of PC2 strategies procession was analyzed with a 2 (sex: male versus female) x 2 (c-fos expression quartile in NAc/PFC: bottom versus top) between-subjects ANOVA. The main effect of sex was significant for both NAc and PFC (NAc: $F(1,28) = 12.87$, $p = 0.001$; PFC: $F(1,28) = 13.47$, $p = 0.001$). F) Causal modeling of the relationship between gene expression level in NAc and PFC and the weight of PC2. The models on top are direct models, indicating that c-fos expression levels in both NAc and PFC are significant predictors of the weights of PC2. The bottom models are the mediation models, in which sex mediated the relationship between neural activity (c-fos expressions in NAc and PFC) and strategy selection (weights of PC2). The arrows are regressions. Paths are labeled with estimated coefficients and significant coefficients are marked by asterisks. The strength of the direct model is greatly reduced and became non-significant when accounted for the mediating effect of sex. This suggests that sex mediated neural measures in explaining strategy selection. Bars ± SEM.

References

1. N. M. Grissom, T. M. Reyes, Let's call the whole thing off: evaluating gender and sex differences in executive function. *Neuropsychopharmacology* (2019) (available at https://www.nature.com/articles/s41386-018-0179-5).

2. C. A. Orsini, B. Setlow, Sex differences in animal models of decision making. *J. Neurosci. Res.* **95**, 260–269 (2017).

3. R. M. Shansky, Are hormones a "female problem" for animal research? *Science*. **364**, 825–826 (2019).

4. J. B. Becker, E. Chartoff, Sex differences in neural mechanisms mediating reward and addiction. *Neuropsychopharmacology*. **44**, 166–183 (2019).

5. Z. Song, M. Kalyani, J. B. Becker, Sex differences in motivated behaviors in animal models. *Curr Opin Behav Sci*. **23**, 98–102 (2018).

6. R. van den Bos, J. Homberg, L. de Visser, A critical review of sex differences in decision-making tasks: Focus on the Iowa Gambling Task. *Behavioural Brain Research*. **238** (2013), pp. 95–108.

7. R. van den Bos, J. Jolles, L. van der Knaap, A. Baars, L. de Visser, Male and female Wistar rats differ in decision-making performance in a rodent version of the Iowa Gambling Task. *Behav. Brain Res.* **234**, 375–379 (2012).

8. J. N. Peak, K. M. Turner, T. H. J. Burne, The effect of developmental vitamin D deficiency in male and female Sprague–Dawley rats on decision-making using a rodent gambling task. *Physiol. Behav.* **138**, 319–324 (2015).

9. C. A. Orsini, M. L. Willis, R. J. Gilbert, J. L. Bizon, B. Setlow, Sex differences in a rat model of risky decision making. *Behav. Neurosci.* **130**, 50–61 (2016).

10. R. M. Shansky, Sex differences in behavioral strategies: avoiding interpretational pitfalls. *Curr. Opin. Neurobiol.* **49**, 95–98 (2018).

11. B. B. Averbeck, V. D. Costa, Motivational neural circuits underlying reinforcement learning. *Nat. Neurosci.* **20**, 505–512 (2017).

12. V. D. Costa, O. Dal Monte, D. R. Lucas, E. A. Murray, B. B. Averbeck, Amygdala and Ventral Striatum Make Distinct Contributions to Reinforcement Learning. *Neuron*. **92**, 505–517 (2016).

13. W. E. Frankenhuis, K. Panchanathan, A. G. Barto, Enriching behavioral ecology with reinforcement learning methods. *Behav. Processes*. **161**, 94–100 (2019).

14. K. M. Rothenhoefer, V. D. Costa, R. Bartolo, R. Vicario-Feliciano, E. A. Murray, B. B.

781 Averbeck, Effects of Ventral Striatum Lesions on Stimulus-Based versus Action-Based
782 Reinforcement Learning. *J. Neurosci.* **37**, 6902–6914 (2017).

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

785 16. A. Izquierdo, C. Aguirre, E. E. Hart, A. Stolyarova, in *Psychiatric Disorders: Methods and*
786 *Protocols*, F. H. Kobeissy, Ed. (Springer New York, New York, NY, 2019;
787 https://doi.org/10.1007/978-1-4939-9554-7_7), pp. 105–119.

788 17. R. B. Ebitz, E. Albarran, T. Moore, Exploration Disrupts Choice-Predictive Signals and
789 Alters Dynamics in Prefrontal Cortex. *Neuron*. **97**, 475 (2018).

790 18. J. M. Pearson, B. Y. Hayden, S. Raghavachari, M. L. Platt, Neurons in posterior cingulate
791 cortex signal exploratory decisions in a dynamic multioption choice task. *Curr. Biol.* **19**,
792 1532–1537 (2009).

793 19. P. H. Rudebeck, E. A. Murray, Balkanizing the primate orbitofrontal cortex: distinct
794 subregions for comparing and contrasting values. *Ann. N. Y. Acad. Sci.* **1239**, 1–13 (2011).

795 20. G. Morris, A. Nevet, D. Arkadir, E. Vaadia, H. Bergman, Midbrain dopamine neurons
796 encode decisions for future action. *Nat. Neurosci.* **9**, 1057–1063 (2006).

797 21. J. M. Pearson, S. R. Heilbronner, D. L. Barack, B. Y. Hayden, M. L. Platt, Posterior
798 cingulate cortex: adapting behavior to a changing world. *Trends Cogn. Sci.* **15**, 143–151
799 (2011).

800 22. M. Pessiglione, B. Seymour, G. Flandin, R. J. Dolan, C. D. Frith, Dopamine-dependent
801 prediction errors underpin reward-seeking behaviour in humans. *Nature*. **442**, 1042–1045
802 (2006).

803 23. M. Steyvers, M. D. Lee, E.-J. Wagenmakers, A Bayesian analysis of human decision-
804 making on bandit problems. *J. Math. Psychol.* **53**, 168–179 (2009).

805 24. S. Zhang, A. J. Yu, in *Advances in Neural Information Processing Systems 26*, C. J. C.
806 Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Q. Weinberger, Eds. (Curran Associates,
807 Inc., 2013; http://papers.nips.cc/paper/5180-forgetful-bayes-and-myopic-planning-human-
808 learning-and-decision-making-in-a-bandit-setting.pdf), pp. 2607–2615.

809 25. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, Cortical substrates for
810 exploratory decisions in humans. *Nature*. **441**, 876–879 (2006).

811 26. H. Kim, J. H. Sul, N. Huh, D. Lee, M. W. Jung, Role of striatum in updating values of
812 chosen actions. *J. Neurosci.* **29**, 14701–14712 (2009).

813 27. M. Ito, K. Doya, Validation of decision-making models and analysis of decision variables in
814 the rat basal ganglia. *J. Neurosci.* **29**, 9861–9874 (2009).

28. J. H. Sul, H. Kim, N. Huh, D. Lee, M. W. Jung, Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron*. **66**, 449–460 (2010).

29. R. Wilson, A. Collins, Ten simple rules for the computational modeling of behavioral data (2019) (available at https://pdfs.semanticscholar.org/91b9/d3ab7532ea24ae70cd726355f25235b1fe8b.pdf).

30. M. Jepma, S. Nieuwenhuis, Pupil Diameter Predicts Changes in the Exploration–Exploitation Trade-off: Evidence for the Adaptive Gain Theory. *J. Cogn. Neurosci.* **23**, 1587–1596 (2011).

31. M. A. Castellano, M. D. Diaz-Palarea, M. Rodriguez, J. Barroso, Lateralization in male rats and dopaminergic system: evidence of right-side population bias. *Physiol. Behav.* **40**, 607–612 (1987).

32. D. Leao Jr, M. Fragoso, P. Ruffino, Regular conditional probability, disintegration of probability and Radon spaces. *Proyecciones*. **23**, 15–29 (2004).

33. A. D. Wyner, A definition of conditional mutual information for arbitrary ensembles. *Information and Control*. **38**, 51–59 (1978).

34. L. Nadel, Ed., in *Encyclopedia of Cognitive Science* (John Wiley & Sons, Ltd, Chichester, 2006; http://doi.wiley.com/10.1002/0470018860.s00585), vol. 63, p. 516.

35. N. Jaworska, A. Chupetlovska-Anastasova, A Review of Multidimensional Scaling (MDS) and its Utility in Various Psychological Domains. *TQMP*. **5**, 1–10 (2009).

36. A. Buja, D. F. Swayne, M. L. Littman, N. Dean, H. Hofmann, L. Chen, Data Visualization With Multidimensional Scaling. *J. Comput. Graph. Stat.* **17**, 444–472 (2008).

37. K. J. Preacher, D. D. Rucker, A. F. Hayes, Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behav. Res.* **42**, 185–227 (2007).

38. M. E. Sobel, Some New Results on Indirect Effects and Their Standard Errors in Covariance Structure Models. *Sociol. Methodol.* **16**, 159–186 (1986).

39. A. M. Zador, A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* **10**, 3770 (2019).

40. R. Sibug, E. Küppers, C. Beyer, S. C. Maxson, C. Pilgrim, I. Reisert, Genotype-dependent sex differentiation of dopaminergic neurons in primary cultures of embryonic mouse brain. *Brain Res. Dev. Brain Res.* **93**, 136–142 (1996).

41. A. P. Arnold, X. Chen, What does the "four core genotypes" mouse model tell us about sex differences in the brain and other tissues? *Front. Neuroendocrinol.* **30**, 1–9 (2009).

42. G. E. Gillies, I. S. Pienaar, S. Vohra, Z. Qamhawi, Sex differences in Parkinson's disease. *Front. Neuroendocrinol.* **35**, 370–384 (2014).

43. M. E. Goldstein, A. W. Tank, L. H. Fossom, R. W. Hamill, Molecular aspects of the regulation of tyrosine hydroxylase by testosterone. *Brain Res. Mol. Brain Res.* **14**, 79–86 (1992).

44. E. S. Calipari, B. Juarez, C. Morel, D. M. Walker, M. E. Cahill, E. Ribeiro, C. Roman-Ortiz, C. Ramakrishnan, K. Deisseroth, M.-H. Han, E. J. Nestler, Dopaminergic dynamics underlying sex-specific cocaine reward. *Nat. Commun.* **8**, 13877 (2017).

45. J. Agrawal, B. Ludwig, B. Roy, Y. Dwivedi, Chronic Testosterone Increases Impulsivity and Influences the Transcriptional Activity of the Alpha-2A Adrenergic Receptor Signaling Pathway in Rat Brain. *Mol. Neurobiol.* **56**, 4061–4071 (2019).

46. L. Bevilacqua, D. Goldman, Genetics of impulsive behaviour. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120380 (2013).

47. J. G. Kie, Optimal Foraging and Risk of Predation: Effects on Behavior and Social Structure in Ungulates. *J. Mammal.* **80**, 1114–1129 (1999).

48. J. B. Becker, E. Chartoff, Sex differences in neural mechanisms mediating reward and addiction. *Neuropsychopharmacology*. **44**, 166–183 (2019).

49. K. A. Uban, J. Rummel, S. B. Floresco, L. A. M. Galea, Estradiol modulates effort-based decision making in female rats. *Neuropsychopharmacology*. **37**, 390–401 (2012).

50. M. M. McCarthy, A. P. Arnold, Reframing sexual differentiation of the brain. *Nat. Neurosci.* **14**, 677–683 (2011).

51. E. K. Miller, J. D. Cohen, An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).

52. R. B. Ebitz, B. J. Sleezer, H. P. Jedema, C. W. Bradberry, B. Y. Hayden, Tonic exploration governs both flexibility and lapses. *PLoS Comput. Biol.* **15**, e1007475 (2019).

53. B. A. Bari, C. D. Grossman, E. E. Lubin, A. E. Rajagopalan, J. I. Cressy, J. Y. Cohen, Stable Representations of Decision Variables for Flexible Behavior. *Neuron*. **103**, 922–933.e7 (2019).

54. A. Mohebi, J. R. Pettibone, A. A. Hamid, J.-M. T. Wong, L. T. Vinson, T. Patriarchi, L. Tian, R. T. Kennedy, J. D. Berke, Dissociable dopamine dynamics for learning and motivation. *Nature*. **570**, 65–70 (2019).

55. C. M. Johnson, H. Peckler, L.-H. Tai, L. Wilbrecht, Rule learning enhances structural plasticity of long-range axons in frontal cortex. *Nat. Commun.* **7**, 10785 (2016).

56. M. Treviño, S. Frey, G. Köhr, Alpha-1 adrenergic receptors gate rapid orientation-specific reduction in visual discrimination. *Cereb. Cortex*. **22**, 2529–2541 (2012).

57. G. T. Prusky, P. W. West, R. M. Douglas, Behavioral assessment of visual acuity in mice

and rats. *Vision Res.* **40**, 2201–2209 (2000).

58. G. Vallortigara, L. J. Rogers, *Behav. Brain Sci.*, in press, doi:10.1017/S0140525X05000105.

59. J. D. Wallis, K. C. Anderson, E. K. Miller, Single neurons in prefrontal cortex encode abstract rules. *Nature*. **411**, 953–956 (2001).

60. M. J. Buckley, F. A. Mansouri, H. Hoda, M. Mahboubi, P. G. F. Browning, S. C. Kwok, A. Phillips, K. Tanaka, Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science*. **325**, 52–58 (2009).

61. D. J. Barraclough, M. L. Conroy, D. Lee, Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neurosci.* **7**, 404–410 (2004).

62. T. J. Bussey, S. P. Wise, E. A. Murray, The role of ventral and orbital prefrontal cortex in conditional visuomotor learning and strategy use in rhesus monkeys (Macaca mulatta). *Behav. Neurosci.* **115**, 971–982 (2001).

63. A. Genovesio, P. J. Brasted, A. R. Mitz, S. P. Wise, Prefrontal cortex activity related to abstract response strategies. *Neuron*. **47**, 307–320 (2005).

64. R. B. Ebitz, J. C. Tu, B. Y. Hayden, Rule adherence warps choice representations and increases decision-making efficiency. *under review*.

65. K. A. Uban, J. Rummel, S. B. Floresco, L. A. M. Galea, Estradiol modulates effort-based decision making in female rats. *Neuropsychopharmacology*. **37**, 390–401 (2012).

66. P. Georgiou, P. Zanos, S. Bhat, J. K. Tracy, I. J. Merchenthaler, M. M. McCarthy, T. D. Gould, Dopamine and Stress System Modulation of Sex Differences in Decision Making. *Neuropsychopharmacology*. **43**, 313–324 (2018).