Divergent strategies for learning in males and females

Cathy S. Chen*[1], R. Becket Ebitz*[2], Sylvia R. Bindas[1], 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

+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

A frequent assumption in value-based decision-making tasks is that agents make decisions based on the feature dimension that reward probabilities vary on. However, in complex, multidimensional environments, stimuli can vary on multiple dimensions at once, meaning that the feature deserving the most credit for outcomes is not always obvious. As a result, individuals may vary in the strategies used to sample stimuli across dimensions, and these strategies may have an unrecognized influence on decision-making. Sex is a proxy for multiple genetic and endocrine influences that can influence decision-making strategies, including how environments are sampled. In this study, we examined the strategies adopted by female and male mice as they learned the value of stimuli that varied in both image and location in a visually-cued two-armed bandit, allowing two possible dimensions to learn about. Female mice acquired the correct image-value associations more quickly than male mice, and they used a fundamentally different strategy to do so. Female mice constrained their decision-space early in learning by preferentially sampling one location over which images varied. Conversely, male strategies were inconsistent, changing frequently and strongly influenced by the immediate experience of stochastic rewards. Individual strategies were related to sex-gated changes in neuronal activation in early learning. Together, we find that in mice, sex is linked with divergent strategies for sampling and learning about the world, revealing substantial unrecognized variability in the approaches implemented during value-based decision-making.

Introduction

Value-based decision-making tasks are used by researchers across species to determine the cognitive and neural mechanisms for reinforcement learning and choice behavior [1–4]. One frequent assumption in value-based decision-making tasks, such as bandit tasks, is that the agents make their decisions based on the feature dimension that the experimenter has designed the reward probabilities to vary on, because the other feature dimensions are uninformative and counterbalanced. However, in complex, multidimensional environments, the feature contributing most to reward outcomes is not always obvious to the individual agent making the choices, and stimuli can simultaneously vary on multiple feature dimensions, such as identity and location, at once [5]. As a result of this complexity, differences in learning and decision making within and between individuals over time could result as much from differences in the strategies employed to sample across the different feature dimensions of stimuli, as they could from more straightforward reinforcement learning processes. Because value-based decision tasks are increasingly used to assess and interpret cognition not only in typical situations, but in conditions relevant to neuropsychiatric disease [6–9], recognizing the diverse strategies that can be employed during value based decision making in typical individuals, and the factors that influence strategy selection, is essential.

Rodents, particularly mice, are increasingly employed to elucidate the fundamental neural mechanisms of value-based decision-making [3,10–14]. One under-recognized strength of mouse models is the ease of running many mice simultaneously and repeatedly, allowing the analysis of individual differences between animals in decision strategies. One potential source of phenotypic variance in value-based decision-making that mice are well-suited to studying is the influence of sex differences. Sex is a proxy for multiple genetic, developmental, and endocrine influences that vary across individual mammals [15–17] and which may be expected to influence how complex, multidimensional environments are sampled and experienced [18–20]. Indeed, sex differences (and gender differences in humans) have been shown in a variety of simple value-based decision-making tasks, but these differences are not always consistent, suggesting the presence of latent variables other than straightforward reinforcement learning properties that may influence how individuals strategically sample their environment and thus learn [21–23]. However, much of the work identifying differences in performance across males and females has used tasks with low

trial counts, choices that vary on only one dimension, and/or confounded rewarding and punishing outcomes that are not well-suited to computationally elucidating the strategies employed during choices. As a result, whether sex differences in value-guided decision making reflect sources of diversity in the strategies employed in sampling multidimensional environments is unknown.

To identify potential sex differences in the strategies employed during value-based decision making, we examined male and female mice as they performed a two-dimensional decision-making task: a two-armed visual bandit[1,2,4,24–28]. While all animals eventually reached the same performance level, female mice learned more rapidly than males and acquired more rewards over the course of learning. Because choice could vary in multiple dimensions in this task[29,30], we were able to consider the possibility that individual animals were adopting different strategies. We found that sex explained a substantial fraction of the individual variability in strategy. Female mice used a systematic strategy where they preferred to choose options in one spatial location, constraining the decision-space and accelerating their learning about image values, even though rewards were counterbalanced across spatial location. Conversely, males simultaneously considered information from both image and spatial dimensions simultaneously, were highly sensitive to the stochastic experience of reward, and changed their own choice strategies frequently. During early learning, gene expression for the neuronal activation marker c-fos in the nucleus accumbens and prefrontal cortex significantly correlated with the extent to which female animals (but not males) used a systematic strategy. These results show that individuals can adopt widely divergent strategies for interacting with the same uncertain world, and that sex is a substantial factor in guiding these strategies.

Results

Age-matched male and female wildtype mice (n=32, 16 per sex, strain B6129SF1/J) were trained to perform a visually-cued two-armed bandit task in touchscreen operant chambers (**Figure 1a**). This task design was similar to those employed in humans and nonhuman primates [1,2,4,24–28,31,32], in contrast to the spatial bandit designs frequently employed with rodents [33–36]. Animals were presented with a repeating set of two different image cues which were associated with different probabilistic reward outcomes (**Figure 1b**).The reward schedule (80%/20%) was

held constant throughout the session. In contrast with spatial bandit designs, here reward contingencies were yoked to image identity, which was randomized with respect to location in the chamber on each trial. This means that the sides (left/right) where image cues appeared were not informative of the reward contingencies. We repeated the task with six different sets of image pairs. Two out of the six tested image pairs were excluded from the study due to extremely high initial preference (>70%) for one image over another. We included four images pairs with equal initial preference for each image and quantified behavioral data in bins of 150 trials for each animal.

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 [4,33,37,38] 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, $p < 0.001$, $\beta 1 = -0.129$; main effect of number of trials, $p < 0.001$, $\beta 2 = -0.017$, 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 [39].

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 strategies (image repeat and spatial repeat) captured the tendency of repeating either the side or the image of the previous trial, regardless of the outcome. Conversely, outcome-dependent strategies followed a win-stay lose-shift policy for either a side or an image, which captured the tendency to only repeat the side or the image of the previous trial when it was rewarded. Fitting the GLM allowed us to estimate how much each of these four strategies was employed within each animal on each bin of trials. We will call this set of weights--the precise pattern of local strategies employed over time--the “global strategy” employed by each individual animal.

Across all animals, we found a global strategy where a specific procession through local strategies was used when learning image pairs (**Figure 2b**). First, animals showed an early tendency towards outcome-independent spatial repeat, giving way to a later interaction between reward outcome and image choice, with outcome-insensitive image repeat (the optimal strategy) increasing in the later stages of testing. To examine whether sex influenced the strength of this strategy procession, we compared the global strategy used by male and female animals. We observed a consistent and pronounced pattern of strategy procession only in females (**Figure 2c**). In contrast, in males, the weight of both image-based strategies simply increased slowly over time (**Figure 2d**), with little change in spatial strategies. Thus, neither a procession of multiple strategies nor a prominent strategy in the early learning stage was observed in male mice.

The sex difference in this strategy selection was intriguing, but it could have been driven by only a few females. Therefore, we next characterized individual variability in strategy via principal component analysis. Here, we estimated the major axes of interindividual variability in strategy, meaning in the unique combinations of the four strategy weight vectors over time and across all animals, regardless of sex. Principal components (PC) 1 and 2 captured the majority of the interindividual variance: 59% of the variability between animals (**Figure 2e**). PC1 reflected a global preference for a side or an image and did not significantly differ between sexes ($p > 0.9$, AUC = 0.43). The mean principal component scores of PC1 for females and males were 0.03 and -0.03, respectively. The mean PC score difference between females and males ($mean_{(F-M)}$) was 0.07 (95% CI = [-1.70, 1.80], $t(30) = 0.08$). Critically, PC2 mirrored the same procession of strategies observed in female, but not male mice (**Figure 2c-d**). This suggests that the extent to which individuals used this procession of strategies explained a large fraction (22%) of the interindividual variability in our animals. The mean PC score of PC2 for females was 0.98 and for males was -0.98. The mean PC score difference between females and males ($mean_{(F-M)}$) was 1.96 (95% CI = [0.87, 3.05], $t(30) = 3.67$). Moreover, knowing the PC2 score (the projection of an individual animal's behavior onto PC2) allowed us to discriminate male and female animals with remarkable accuracy (reciever operating characteristic analysis, AUC = 0.86, significant discrimination: $p < 0.001$). No otherPCs differed between sexes ($p > 0.4$, AUC < 0.6). Together, these results suggest that the choice of what strategy to follow explains substantial individual

variability in multidimensional decision-making, and that differences in strategy can depend on sex.

There are two possible explanations why females consistently implemented the strategy procession captured by PC2. One hypothesis is that this early side bias reflected an *energy saving* strategy that saved time and/or effort by just repeating the same side. Alternatively, this early side bias could be an *cognitively effortful* strategy to constrain decision-making to one 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, $\beta3 = 0.03$, $p = 0.0007$) and significantly slower than males across all trials (GLM, main effect of sex, $\beta1 = -0.62$, $p < 0.0001$). 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, $\beta2 = -0.04$, $p<0.0001$, 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 [40,41], 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, $\beta 1 = -1.47$, $p = 0.11$, 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, $\beta 1 = 4.24$, $p < 0.0001$, see equation 1 in Methods) and across multiple image pairs (**Figure 4f**, GLM, main effect of sex, $\beta 1 = 4.54$, $p = 0.047$, see equation 1 in Methods). Overall, the variability in choices decreased across time, as the divergent strategies used by individual animals started to

284 converge to the optimal strategy in this task, which is to choose the high-value image
285 consistently (GLM, main effect of number of trials, within sex between suject: $\beta 2 = -0.78$, $p <$
286 0.0001; within subject between bins: $\beta 2 = -0.359$, $p < 0.0001$). Together, these results suggest
287 that individual males displayed divergent choice patterns and were changing between complex
288 strategies over time and the repetition of the same task, while females largely adopted a shared,
289 systematic approach to learning.
290
291 To visualize animals' patterns of choices expressed in the probability vectors, we used
292 multidimensional scaling (MDS) [42–44] to reduce the dimensionality of strategy space,
293 allowing us to project the high-dimensional "strategy path" throughout learning onto a 2
294 dimensional space. This allows us to easily visualize the similarity between patterns of choice
295 across animals over time and across repetitions. Each color path represents a strategy path used
296 in a different repetition of the task (4 repetitions in total). **Figure 4g** shows examples of strategy
297 paths of males and females. The optimal strategy in this bandit task, which is to choose the high
298 value image consistently regardless of the outcome, is represented by a star in the low
299 dimensional space. Both males and females "strategy path" showed gradual approximation to the
300 optimal strategy over time. Consistent with the quantification described above, the strategy path
301 of males are visibly more variable and different across repetitions of the task, whereas the
302 strategy path of females were more unified and consistent across repetitions.
303
304 **Sex mediated the ability of neuronal activity to explain strategy selection**
305 The ability to learn and perform bandit tasks is highly sensitive to alterations in corticolimbic
306 structures. However, it remains unclear how alterations in these structures predict choice
307 strategy, much less sex differences in choice strategy. To address this question, we examined
308 neuronal activity in several corticolimbic brain regions by examining the expression of *c-fos,* an
309 immediate early gene often used as a marker of neuronal activation. The animals from the
310 previous figures were sacrificed after the second day of a new, final image-reward pairing (each
311 animal having completed 400-500 total trials), corresponding to when the female side bias was
312 greatest. We compared mRNA expression level for *c-fos* across five brain regions, including
313 nucleus accumbens (NAc), dorsal medial striatum (DMS), amygdala (AMY), hippocampus
314 (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 regions 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 [45,46] 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

346 expression in NAc/PFC confirmed that neural activation is a significant predicor of the mediator
347 sex (NAc: $\alpha = 0.20$, $p = 0.024$; PFC: $\alpha = 0.26$, $p < 0.004$, see equation 7 in Methods). When we
348 regressed strategy selection on both the mediator variable (sex) and independent variable (neural
349 activation in NAc/PFC), the result showed that the mediator sex was a significant predictor of
350 strategy selection (NAc: $\beta1 = 1.66$, $p = 0.008$; PFC: $\beta1 = 1.64$, $p = 0.012$), and the strength of the
351 direct model is now greatly reduced and became non-significant when accounted for the
352 mediating effect of sex (NAc: $\beta' = 0.38$, $p = 0.20$; PFC: $\beta' = 0.33$, $p = 0.31$). The Sobel (1982)
353 first-order test was used to assess the presence of mediation [46]. The indirect effect was
354 calculated as the product of coefficients and was significant for both NAc and PFC (NAc: $\alpha\beta' =$
355 0.34, $z = 1.836$, $p < 0.039$; PFC: $\alpha\beta' = 0.42$, $z = 2.035$, $p < 0.026$). Together, these results
356 suggest that the relationship between PFC and NAc c-fos and PC2 differed, depending on the
357 animals' sex. This suggests that sex-linked mechanisms gate the relationship between these
358 circuits and strategic decision-making and highlight these regions as promising targets for future
359 studies looking at the effects of sex on the neural circuits responsible for implementing strategic
360 learning.
361
362 **Discussion**
363
364 By training male and female mice on a stochastic two-dimensional decision-making task, we
365 were able to evoke a range of problem solving strategies across individuals. In this task, each cue
366 has two dimensions - the identity of the image and the location of the image. Animals had to
367 explore the reward value associated with both cue dimensions to determine which were most
368 predictive of reward. Although both male and female mice eventually learned the right strategy,
369 choosing the high-value image, female mice learned faster. The richness of this task allowed us
370 to uncover sex differences in *how* the animals achieved the associations across time. We
371 discovered that female mice adopted a consistent and systematic approach where they processed
372 through different strategies over time. Early in learning, they constrained their search space by
373 only sampling the outcomes of images on one side (left or right). This approach, which occurred
374 when animals were most uncertain about the best choice, reduced the number of dimensions they
375 were learning about and permitted more rapid acquisition of the image-value association. In
376 contrast, males employed a strategy of decisions that seemed to combine both image and spatial

377 location, changed frequently, and was strongly influenced by the immediate prior experience of
378 reinforcement. While both sexes eventually reached equivalent levels of performance, our data
379 reveal that the strategic paths journeys individual animals take to get there can vary dramatically,
380 implicating the potential for wide divergence in neural circuit mechanisms in normative decision
381 making.
382
383 Sequential decision-making and learning in rodents is often studied with spatial bandit tasks, in
384 which reward probabilities are linked to left and right levers or sides that are visually identical
385 [1,4,11,13,33,47]. In these spatial bandit tasks, side bias in choice has sometimes been reported
386 as a behavioral artifact and animals displaying such bias were often excluded from experiments
387 [48–50]. However, in the current task, both the side and the identity of the image cues could have
388 been informative of the reward probabilities (although they were not). In principle, animals could
389 simultaneously sample both dimensions to learn side values and image values at the same time.
390 However, in practice, it appears that the early side bias in female mice "jump-started" their
391 learning by controlling for space while exploring choice-outcome values of the images, which in
392 this task happened to be the more informative dimension of reward. Intriguingly, this suggests
393 that females were covertly learning about the correct cue dimension while behaviorally selecting
394 the wrong item, and were able to convert this to successful learning due to the stability of the
395 task structure. This view suggests that we should be able to design tasks that prevent the
396 successful use of this strategy, and which might therefore shift the presentation of the sex
397 difference in decision-making.
398
399 Our data implicate the prefrontal cortex (PFC) and nucleus accumbens (NAc), part of the ventral
400 striatum, in the differences in strategy between males and females. These regions have been
401 widely implicated in reward-guided decision making, but so have the other regions we tested for
402 which we didn't find a significant relationship to behavior [2,29,51]. One possibility is that the
403 PFC and accumbens are particularly engaged in strategic decision-making. This resonates with
404 previous studies that have implicated the PFC in implementing strategies and rule-guided
405 behaviors [47,52–57] and the NAc in selecting and implementing learning strategies [29,30].
406 Implementing different strategies produces changes in how different choice dimensions are
407 represented in the PFC and NAc [58], and lesions in the NAc can drive animals towards a low-

408 dimensional action-based strategy or prevent animals from switching between strategies [2,30].
409 The PFC is also sensitive to gonadal hormones during risky decision making [59], and
410 dopaminergic function in the accumbens regulates risky decision making in a sex-specific
411 manner [60], perhaps due to sex differences in dopamine neurons [61]. Here, the relationship
412 between both PFC and NAc and strategy use was mediated by sex, suggesting that whatever the
413 relationship between these regions and strategic decision-making, it is likely to be sex-biased.
414
415 One fundamental unanswered question is *why* females as a group employed a highly similar and
416 consistent strategy. Zador (2019) recently proposed that much of animal behavior is not dictated
417 by supervised or unsupervised learning algorithms, but are largely shaped by biological
418 constraints [62]. Energy-conserving and habitual behaviors are prevalent in female animals,
419 including during foraging [22,63,64]. The biological constraints and organization imposed by the
420 multiple mechanisms of sexual differentiation are known to drive a tuning of the circuits
421 important for reward-guided decisions [61,65]. Exogenously applied gonadal steroid hormones
422 are known to modulate cost/benefit decision-making. While testosterone promotes effort
423 expenditure and impulsive behavior [66–68], estradiol reduces high-effort choices [69–71]. In
424 addition, sex chromosomes exert independent influences on reward-guided behaviors [15] with
425 elevated habit in XX carriers and increased effort in XY carriers, regardless of gonadal hormone
426 status [72,73]. Of course, these impacts of sexual differentiation are graded, rather than
427 dichotomous across the sexes. Indeed, here we found that a small number of males showed some
428 tendency to use the female strategy, implicating graded mechanisms of early-life and/or pubertal
429 masculinization in the degree of adoption of sex-biased exploratory strategies we observed here.
430 An intriguing possibility is that the unified, consistent, and systematic strategy we observed in
431 female mice, as well as the volatile and diverse strategy we observed in male mice in the same
432 task, may emerge from evolved sex-biased strategies for foraging in the wild that were critical to
433 survival for the species as a whole.

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 occured 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

Pretraining. animals were exposed daily to a 30-min session of initial touch training, during which a blank white square (cue) was presented on one side of the touchscreen, counterbalancing left and right between trials. This schedule provided free reinforcement every 30 seconds, during which the cue was on. If animals touched the cue during this period, a reward three times the size of the regular reward was dispensed (840 ms). This led to rapid acquisition. Following this, animals were exposed daily to a 30-min session of must touch training. This schedule followed the same procedure as the initial touch training, but free reinforcers were terminated and animals were required to nose poke the image in order to obtain a regular reward (7-uL, 280 ms).

Deterministic pairwise discrimination training. Animals were exposed to 10 days of pairwise discrimination training, during which animals were presented with two highly discriminable image cues ("marbles" and "fan"). One image was always rewarded and the other one was not. Within each session, animals completed either 250 trials or spent a maximum of two hours in the operant chamber (typically these mice completed ~200 trials/day).

Two-armed bandit task. Animals were trained to perform a two-arm visual bandit task in the touchscreen operant chamber. Each trial, animals were presented with a repeating set of two different images on the left and right side of screen, counterbalancing left and right across the

session. Nose poke to one of the displayed images on the touchscreen was required to register a response. Nose poke on one image triggered a reward 80% of the time (high payoff image), whereas the other image was only reinforced 20% of the time (low payoff image). Following the reward collection, which was registered as entry and exit of the feeder hole, the magazine would 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 [39], calculated for each mouse according to the following formula:

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

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

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

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

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

Meaning that the current position of image 1 determined the sign of the side bias term. This model was fit individually to each bin of 150 trials, within each animal and image pair, via cross-entropy minimization with a regularization term (L2/ridge regression).

Principal component analysis. In order to determine how decision-making strategies differed across animals and bins, we looked for the major dimensions of interindividual variability in decision-making strategies. To do this, we took advantage of the fact that the coefficients of the generalized linear model provided a simplified description of how decision-making depended on image, side, and outcome for each subject in each bin. Because the generalized logistic regression model estimated 4 terms per image pair and there were 23 independent bins per image pair, this meant that each animals' behavior for a given image pair could be described as a 4*23 by 1 dimensional vector. We then used principal component analysis to identify the linear combinations of model parameters that explained the most variance across subjects and repetitions of image pairs (across 32 (animals) x 4 (repetitions) = 128 total strategy vectors). The first two principal components, which explained the majority of the variance (59%), are illustrated in Figure 2e.

Conditional mutual information and model-free analyses. To account for idiosyncratic strategies, which could vary across animals or image pairs, we used a model-free approach to quantify the extent to which behavior was structured without making strong assumptions about what form this structure might take. We quantified the extent to which choice history was informative about current choices as the conditional mutual information between the current choice (C) and the last 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.

Mediation Analysis. First we used a direct model and regressed c-fos expression of either NAc or PFC on weights of PC2. When assessing a mediation effect, three regression models are examined:

Model 1 (direct):

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

Model 2 (mediation):

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

Model 3 (indirect)

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

In these models, γ1, γ2, and γ3 represent the intercepts for each model, while ε1, ε2, and ε3 represent the error term. β denotes the relationship between dependent variable (PC2 weights) and independent variable (NAc c-fos expression) in the first model, and β'denotes the same relationship in the third model. α represents the relationship between independent variable (NAc c-fos expression) and mediator (sex) in the second model. The mediation effect is calculated using the product of coefficients (αβ1). The Sobel test is used to determine whether the mediation effect is statistically significant [46].

Reinforcement Learning Model.

We considered a basic reinforcement learning model, following the Rescorla Wagner rule. In this model, subjects first learn the expected value of each image based on the history of its previous outcome value Q and use these Q values to decide what to do next. The expected value of arm *k* on the *t* th trial, Q_t^k, is updated based on the reward outcome of each trial:

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

In each trial, $r_t - Q_t^k$ captures the reward prediction error (RPE), which is the diffe-rence between expected outcome and the actual outcome. The parameter *a* is the learning rate, which determines the rate of updating RPE. Action selection was performed based on a Softmax probability distribution:

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

where inverse temperature *b* determines the level of random exploration.

Acknowledgements

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

Figures and Figure Captions

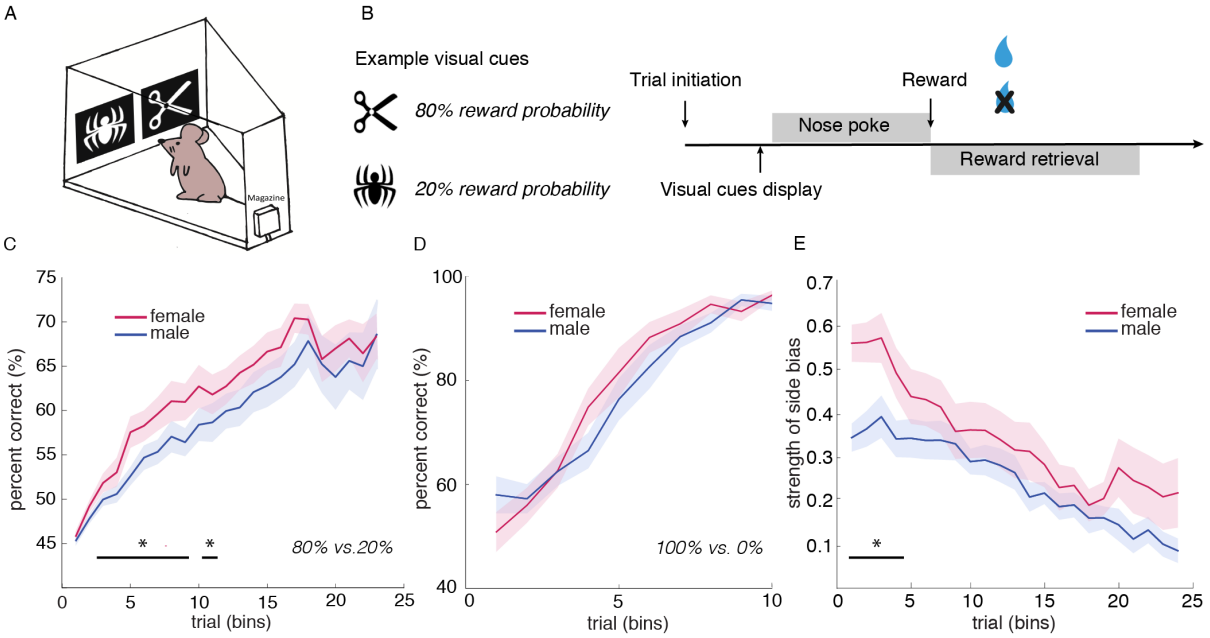

Figure 1. **Females showed accelerated acquisition of the high reward probability image in a stochastic two-armed visual bandit task.** A) Schematic of the mouse touch-screen operant chamber used in our task. B) Schematic of two-armed Visual Bandit task. Images varied between the two locations across 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 $p < 0.05$. Graphs depict mean ± SEM.

Figure 2

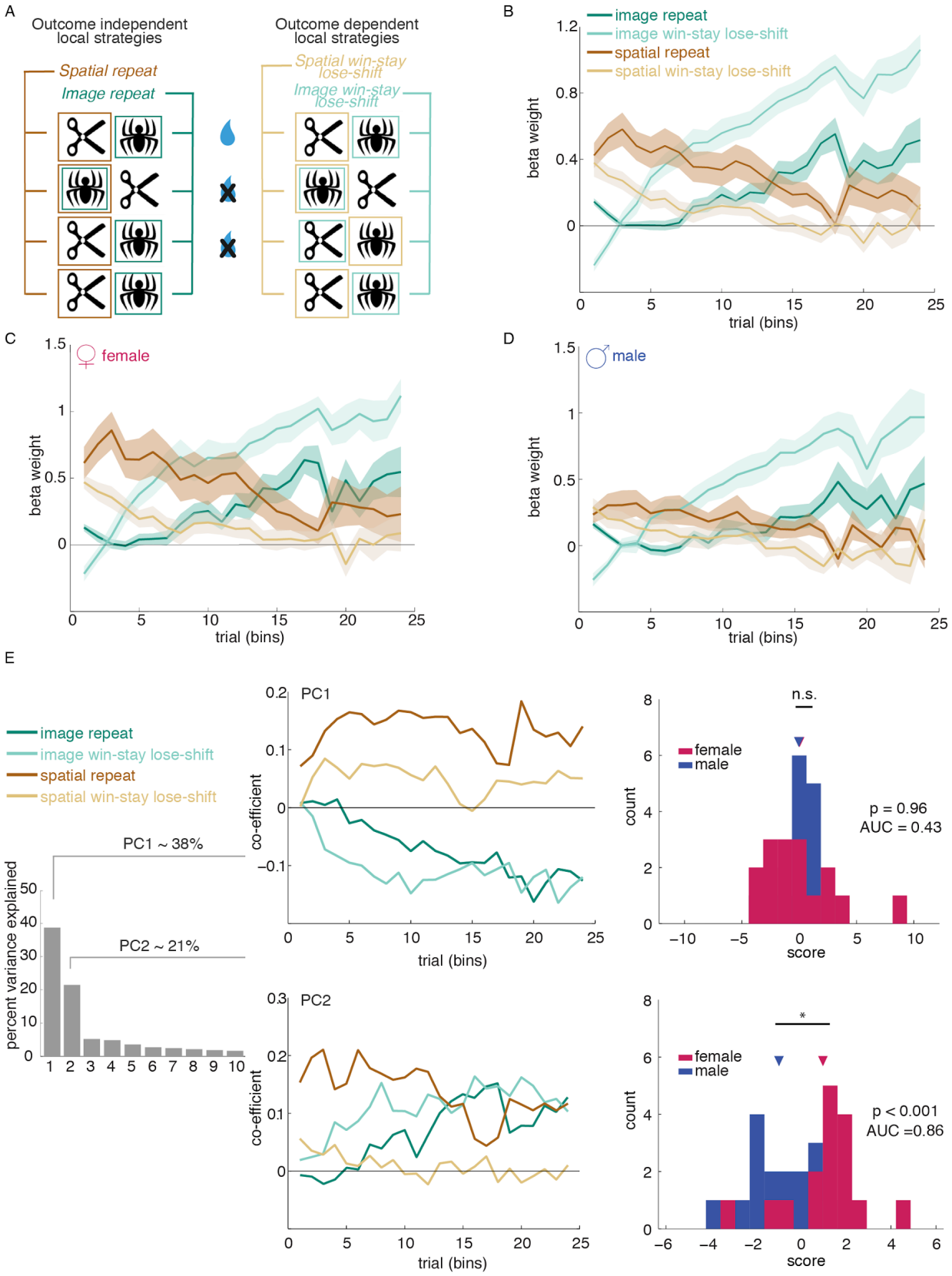
A
Outcome independent local strategies
Spatial repeat
Image repeat
Outcome dependent local strategies
Spatial win-stay lose-shift
Image win-stay lose-shift
B
image repeat
image win-stay lose-shift
spatial repeat
spatial win-stay lose-shift
beta weight
trial (bins)
C
female
beta weight
trial (bins)
D
male
beta weight
trial (bins)
E
image repeat
image win-stay lose-shift
spatial repeat
spatial win-stay lose-shift
PC1 ~ 38%
PC2 ~ 21%
percent variance explained
PC1
co-efficient
trial (bins)
n.s.
female
male
count
score
p = 0.96
AUC = 0.43
PC2
co-efficient
trial (bins)
*
female
male
count
score
p < 0.001
AUC =0.86

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. Graphs depict mean ± SEM.

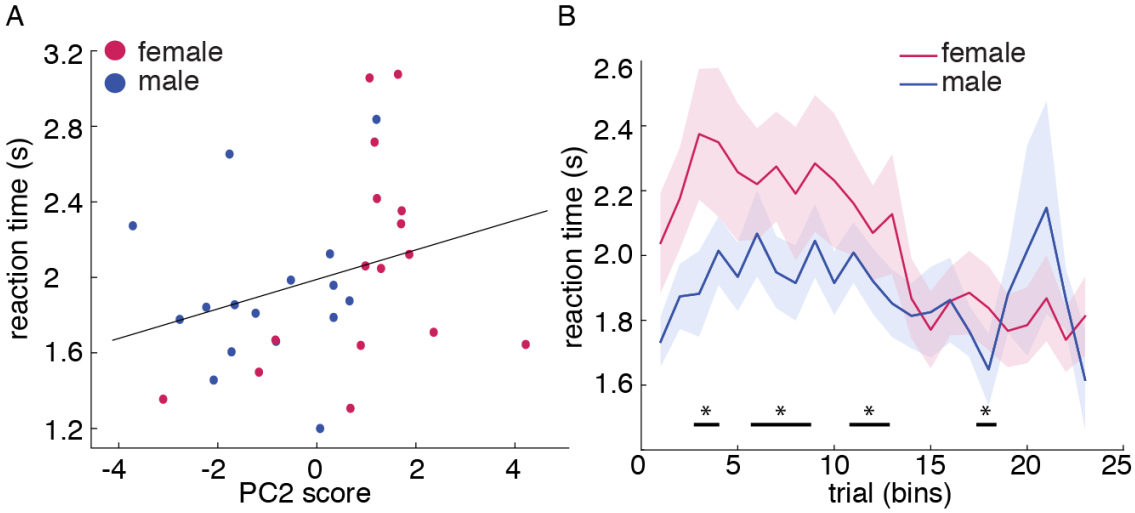

Figure 3. The side-to-image strategy procession captured by Principal Component (PC) 2 is a cognitively demanding strategy but not a time-saving strategy. A) Correlation analyses revealed a significant positive correlation between PC2 scores and reaction time. The decision-making time was longer within animals primarily used PC2 strategy. B) Predominantly using PC2, females responded slower during early learning (GLM, interaction term, $\beta 3 = 0.03$, $p = 0.0007$). Note that, the slow reaction time during early learning in females matches with the time period (bin 1-10) during which females relied on side-bias "heuristics" for decision-making. Data shown as bins of 150 trials. * indicates $p < 0.05$. Graph B depicts mean ± SEM.

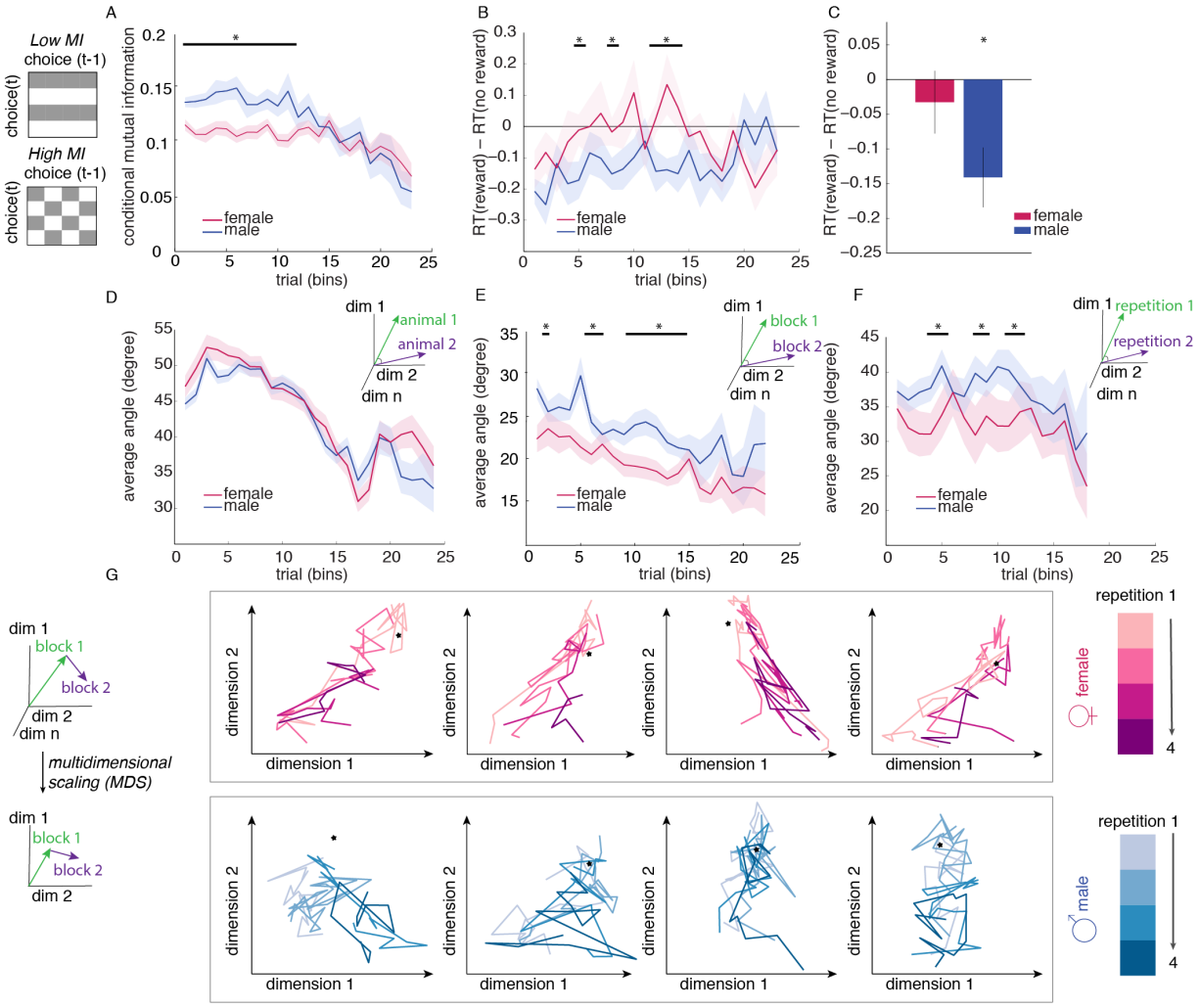

Figure 4. Male mice were more likely to differ from themselves over time, with choice patterns dependent on past outcomes. We expressed each animal's history-dependent choice pattern as an 32-dimensional vector of joint probabilities and measured the angle between vectors, which is proportional to the 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 $p < 0.05$. Graphs depict mean ± SEM.

706 **Figure 5**

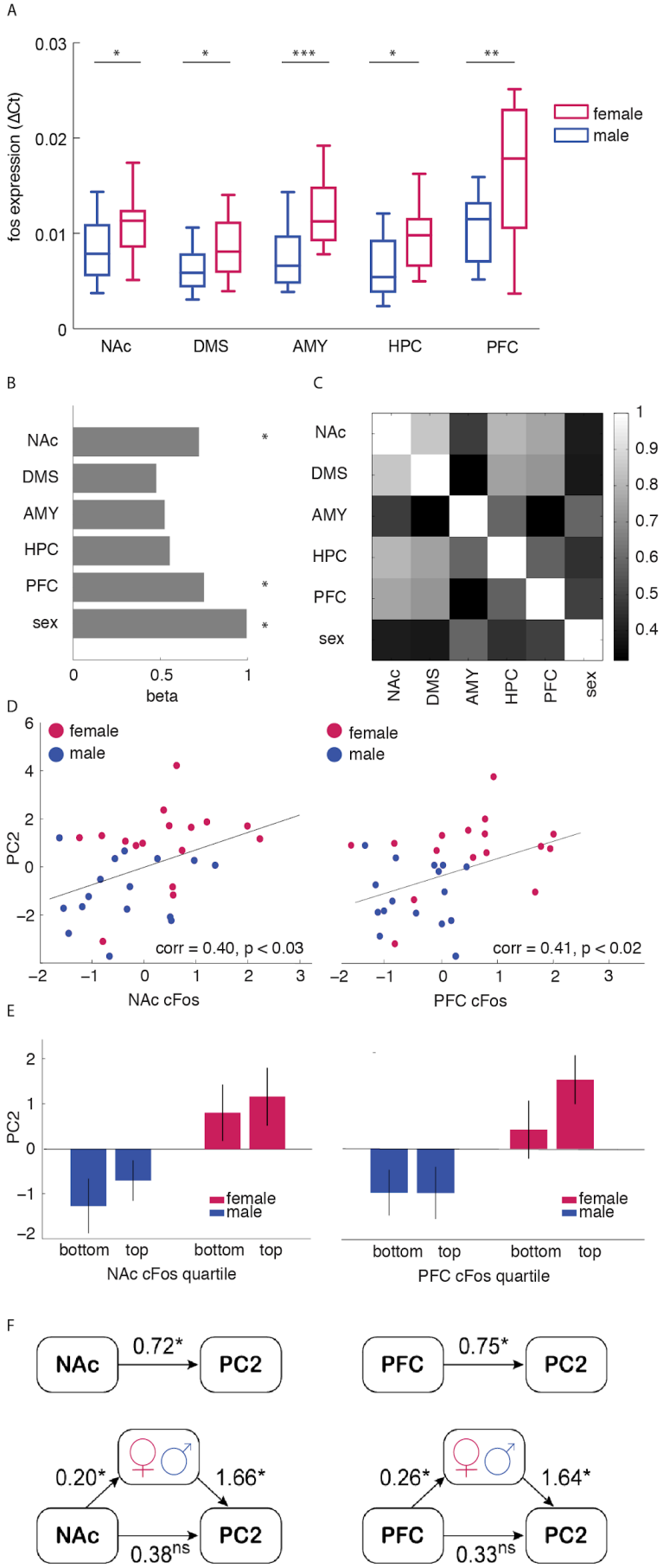
A
fos expression (ΔCt)
0.03
0.02
0.01
0
*
*
***
*
**
female
male
NAc
DMS
AMY
HPC
PFC
B
NAc
DMS
AMY
HPC
PFC
sex
*
*
*
0
0.5
1
beta
C
NAc
DMS
AMY
HPC
PFC
sex
1
0.9
0.8
0.7
0.6
0.5
0.4
D
female
male
PC2
corr = 0.40, p < 0.03
NAc cFos
corr = 0.41, p < 0.02
PFC cFos
E
PC2
female
male
bottom
top
NAc cFos quartile
PFC cFos quartile
F
NAc
0.72*
PC2
PFC
0.75*
0.20*
1.66*
0.38ns
0.26*
1.64*
0.33ns

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. Graphs depict mean ± SEM.

References

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

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

3. Parker, N.F., Cameron, C.M., Taliaferro, J.P., Lee, J., Choi, J.Y., Davidson, T.J., Daw, N.D., and Witten, I.B. (2016). Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat. Neurosci. *19*, 845–854.

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

5. Bartolo, R., Saunders, R.C., Mitz, A., and Averbeck, B.B. (2019). Dimensionality, information and learning in prefrontal cortex. bioRxiv, 823377. Available at: https://www.biorxiv.org/content/10.1101/823377v1.full [Accessed February 25, 2020].

6. Grissom, N.M., McKee, S.E., Schoch, H., Bowman, N., Havekes, R., O'Brien, W.T., Mahrt, E., Siegel, S., Commons, K., Portfors, C., *et al.* (2018). Male-specific deficits in natural reward learning in a mouse model of neurodevelopmental disorders. Mol. Psychiatry *23*, 544–555.

7. Solomon, M., Smith, A.C., Frank, M.J., Ly, S., and Carter, C.S. (2011). Probabilistic reinforcement learning in adults with autism spectrum disorders. Autism Res. *4*, 109–120.

8. Grissom, N., McKee, S., Schoch, H., Bowman, N., Havekes, R., Nickl-Jockschat, T., Reyes, T., and Abel, T. (2015). Male-Specific Reward Learning Deficits in a Mouse Model of Autism. In NEUROPSYCHOPHARMACOLOGY (NATURE PUBLISHING GROUP MACMILLAN BUILDING, 4 CRINAN ST, LONDON N1 9XW, ENGLAND), pp. S293–S293.

9. Kim, H., Lee, D., Shin, Y.-M., and Chey, J. (2007). Impaired strategic decision making in schizophrenia. Brain Res. *1180*, 90–100.

10. Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H.J., Ornelas, S., Koay, S.A., Thiberge, S.Y., Daw, N.D., Tank, D.W., *et al.* (2019). Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. Nature, 1.

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

12. Alabi, O.O., Fortunato, M.P., and Fuccillo, M.V. (2019). Behavioral Paradigms to Probe Individual Mouse Differences in Value-Based Decision Making. Front. Neurosci. *13*, 50.

765 13. Bari, B.A., Grossman, C.D., Lubin, E.E., Rajagopalan, A.E., Cressy, J.I., and Cohen, J.Y.
766 (2019). Stable Representations of Decision Variables for Flexible Behavior. Neuron *103*,
767 922–933.e7.

768 14. Cohen, J.Y. (2015). Dopamine and serotonin signals for reward across time scales. Science
769 *350*, 47.

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

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

774 17. Shansky, R.M. (2019). Are hormones a "female problem" for animal research? Science *364*,
775 825–826.

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

778 19. Gruene, T.M., Flick, K., Stefano, A., Shea, S.D., and Shansky, R.M. (2015). Sexually
779 divergent expression of active and passive conditioned fear responses in rats. Elife *4*.
780 Available at: http://dx.doi.org/10.7554/eLife.11352.

781 20. McCarthy, M.M., Arnold, A.P., Ball, G.F., Blaustein, J.D., and De Vries, G.J. (2012). Sex
782 differences in the brain: the not so inconvenient truth. J. Neurosci. *32*, 2241–2247.

783 21. Grissom, N.M., and Reyes, T.M. (2018). Let's call the whole thing off: evaluating gender
784 and sex differences in executive function. Neuropsychopharmacology. Available at:
785 http://dx.doi.org/10.1038/s41386-018-0179-5.

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

788 23. van den Bos, R., Homberg, J., and de Visser, L. (2013). A critical review of sex differences
789 in decision-making tasks: Focus on the Iowa Gambling Task. Behavioural Brain Research
790 *238*, 95–108. Available at: http://dx.doi.org/10.1016/j.bbr.2012.10.002.

791 24. Izquierdo, A., Aguirre, C., Hart, E.E., and Stolyarova, A. (2019). Rodent Models of
792 Adaptive Value Learning and Decision-Making. In Psychiatric Disorders: Methods and
793 Protocols, F. H. Kobeissy, ed. (New York, NY: Springer New York), pp. 105–119.

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

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

798 27. Pearson, J.M., Heilbronner, S.R., Barack, D.L., Hayden, B.Y., and Platt, M.L. (2011).

Posterior cingulate cortex: adapting behavior to a changing world. Trends Cogn. Sci. *15*, 143–151.

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

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

30. Rothenhoefer, K.M., Costa, V.D., Bartolo, R., Vicario-Feliciano, R., Murray, E.A., and Averbeck, B.B. (2017). Effects of Ventral Striatum Lesions on Stimulus-Based versus Action-Based Reinforcement Learning. J. Neurosci. *37*, 6902–6914.

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

32. Zhang, S., and Yu, A.J. (2013). Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting. In Advances in Neural Information Processing Systems 26, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds. (Curran Associates, Inc.), pp. 2607–2615.

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

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

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

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

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

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

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

40. Leao, D., Jr, Fragoso, M., and Ruffino, P. (2004). Regular conditional probability,

disintegration of probability and Radon spaces. Proyecciones *23*, 15–29.

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

42. Nadel, L. ed. (2006). Multidimensional Scaling. In Encyclopedia of Cognitive Science (Chichester: John Wiley & Sons, Ltd), p. 516.

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

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

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

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

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

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

49. Prusky, G.T., West, P.W., and Douglas, R.M. (2000). Behavioral assessment of visual acuity in mice and rats. Vision Res. *40*, 2201–2209.

50. Vallortigara, G., and Rogers, L.J. (2005). Survival with an asymmetrical brain: advantages and disadvantages of cerebral lateralization. Behav. Brain Sci. *28*, 575–89; discussion 589–633.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

891 66. Goldstein, M.E., Tank, A.W., Fossom, L.H., and Hamill, R.W. (1992). Molecular aspects of
892 the regulation of tyrosine hydroxylase by testosterone. Brain Res. Mol. Brain Res. *14*, 79–
893 86.

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

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

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

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

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

905 72. Seu, E., Groman, S.M., Arnold, A.P., and Jentsch, J.D. (2014). Sex chromosome
906 complement influences operant responding for a palatable food in mice. Genes Brain
907 Behav. *13*, 527–534.

908 73. Quinn, J.J., Hitchcott, P.K., Umeda, E.A., Arnold, A.P., and Taylor, J.R. (2007). Sex
909 chromosome complement regulates habit formation. Nat. Neurosci. *10*, 1398–1400.