

1 **Touchscreen response precision is sensitive to the explore/exploit tradeoff**

2

3 **Authors:** Dana Mueller¹, Erin Giglio¹, Cathy S. Chen¹, Aspen Holm¹, R. Becket Ebitz², Nicola M.
4 Grissom¹

5

6

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

8 ² Department of Neurosciences, Université de Montréal, Quebec, Canada

9

10

11 **+to whom correspondence should be addressed:**

12 Nicola Grissom

13 Department of Psychology

14 University of Minnesota, Minneapolis, MN

15 75 East River Road

16 Minneapolis, MN 55454

17

18 Email: ngrissom@umn.edu

19 **Abstract**

20 The explore/exploit tradeoff is a fundamental property of choice selection during reward-guided
21 decision making. In perceptual decision making, higher certainty decisions are more motorically
22 precise, even when the decision does not require motor accuracy. However, while we can
23 parametrically control uncertainty in perceptual tasks, we do not know what variables - if any -
24 shape motor precision and reflect subjective certainty during reward-guided decision making.
25 Touchscreens are increasingly used across species to measure choice, but provide no tactile
26 feedback on whether an action is precise or not, and therefore provide a valuable opportunity to
27 determine whether actions differ in precision due to explore/exploit state, reward, or individual
28 variables. We find all three of these factors exert independent drives towards increased
29 precision. During exploit states, successive touches to the same choice are closer together than
30 those made in an explore state, consistent with exploit states reflecting higher certainty and/or
31 motor stereotypy in responding. However, exploit decisions might be expected to be rewarded
32 more frequently than explore decisions. We find that exploit choice precision is increased
33 independently of a separate increase in precision due to immediate past reward, suggesting
34 multiple mechanisms regulating choice precision. Finally, we see evidence that male mice in
35 general are less precise in their interactions with the touchscreen than females, even when
36 exploiting a choice. These results suggest that as exploit behavior emerges in reward-guided
37 decision making, individuals become more motorically precise reflecting increased certainty,
38 even when decision choice does not require additional motor accuracy, but this is influenced by
39 individual differences and prior reward. These data uncover the hidden potential for touchscreen
40 tasks in any species to uncover the latent neural states that unite cognition and movement.

41 **Key Words**

42 Reinforcement Learning, Bandit, Sex Differences, Touchscreen, Hidden Markov model (HMM)

43 **Introduction**

44 Sequential reward-guided decision making tasks, such as multi-armed bandit tasks, are well
45 known to engage explore/exploit tradeoffs (Addicott et al., 2017; Chen et al., 2023a, 2021b;
46 Ebitz et al., 2019, 2018; Stephens, 2008; Wyatt et al., 2023). Across species, exploration
47 represents periods of variable choice selection and heightened learning about the environment,
48 relative to exploit behaviors, which show consistent choice selection that is less sensitive to trial-
49 to-trial feedback (Badre et al., 2012; Cavanagh et al., 2012; Daw et al., 2005; Frank and
50 Fossella, 2011; Ting et al., 2023; Trudel et al., 2021). Explore/exploit tradeoffs therefore reveal
51 that superficially similar choice behaviors (for example left vs right choice) can be driven by
52 highly distinct neural states (Ebitz et al., 2020, 2019, 2018; Wang et al., 2023; Wyatt et al.,
53 2023), likely reflecting differences in the certainty of choices. However, certainty in reward-
54 guided tasks is individual and subjective, and we do not have good ways of measuring it without
55 self report.

56
57 Perceptual decision making tasks reveal that higher certainty decisions are more motorically
58 precise, even when the decision does not require motor accuracy (Follman et al., 2023; Palser
59 et al., 2018; Sanchez et al., 2024; Wolpert and Landy, 2012). However, while we can
60 parametrically control uncertainty in perceptual tasks, we do not know what variables--if any--
61 shape motor precision during other forms of decision-making. This is an especially significant
62 omission in the case of reward-guided decision-making because precision could be influenced
63 either by prior rewards (which increase certainty about the correct action) or by decision-making
64 states (which may or may not increase certainty). We do have reason to believe that
65 explore/exploit states reflect differences in certainty - for example, exploitative choices are faster
66 than exploratory ones (Addicott et al., 2017; Chen et al., 2023a, 2021a, 2021b; Ebitz et al.,
67 2018; Hassall et al., 2013; Laureiro-Martinez et al., 2010; Walker et al., 2022; Wershbaile and
68 Pleskac, 2010). Although explore and exploit strategies are defined at the broadest level by the
69 options chosen in a decision making task, these findings strongly imply that explore/exploit
70 balance is also reflected in the fine-grained execution of the task.

71
72 Touchscreen operant chambers in animal models offer a powerful and novel approach in
73 exploring the kinetics of a choice response by logging the precise coordinates and timing of
74 each choice that is made on the screen, across thousands of choices. Screens by default offer
75 no immediate, tactile feedback about choice accuracy, requiring that longer trial-and-error
76 processes influence touch similarity. Pigeons and other birds have shown an awareness of
77 spatial location of touches on touchscreens and make minute adjustments of touches as the
78 task evolves, suggesting that the same might be evident for rodents (Capshew, 1993; Goodale,
79 1983; Jager and Zeigler, 1991; Peterson, 2004; Skinner, 1960; Spetch et al., 1992). We took
80 advantage of this rich but underutilized data to analyze the location of decision touches across
81 sexes from explore/exploit data in mice we have previously published (Chen et al., 2021b),
82 asking if explore/exploit balance governed how similar choice touches were from one trial to the
83 next. We found that actions become more precise in exploit state behavior compared to explore
84 state. This effect was independent of a similar effect of reward on touch location, suggesting
85 parallel mechanisms by which explore/exploit state and prior outcomes influence the precision

86 of the next action execution. Because male and female mice employ different strategies in the
87 two-arm restless bandit task, we tested whether the precision of choices to the screen was
88 modulated by sex, and found that actions were more precise in females compared to males,
89 also independent of the impact of explore/exploit state and reward experience, suggesting
90 individual differences regulating action precision over and above other cognitive features of the
91 task. Overall, this novel analysis capitalizes on the hidden potential for touchscreens to measure
92 not only choice behaviors but the motor actions that generate them, informing the neural states
93 that unite movement and cognition.

94 **Results**

95 To understand how actions in the chamber are influenced by internal states in the animals and
96 external events, we took advantage of a previously collected dataset examining sex differences
97 in explore/exploit balance in mice in a touchscreen bandit task. Decision making data from the
98 experiments analyzed here were originally shared in Chen et al. 2021. These data were
99 collected from age-matched male and female wild-type mice ($n = 32$, 16 per sex, strain
100 B6129SF1/J). Mice were trained in a two-arm spatial restless bandit task (**Figure 1a, 1c**) in a
101 trapezoidal shaped touchscreen operant chamber. In this bandit task the probability of reward of
102 each left and right choice changes independently and randomly of the other, with a 10% chance
103 of probability change on each trial (**Figure 1c**: example probability walk). The unpredictability of
104 this task encourages mice to continually learn and survey their choices, exploring to find the
105 best option and exploiting a good rewarding option across a 300 trial session. Explore and
106 exploit trials were labeled using a Hidden Markov model (HMM) approach (Chen et al., 2021b;
107 Ebitz et al., 2018) where a mouse could either explore, exploit left choice, or exploit right choice
108 (**Figure 1c**). Each trial nosepoke response on the touchscreen can therefore be identified as an
109 explore or exploit choice (**Figure 1b**).

110 **Exploit states and female sex are associated with reduced action variability**

111 Using previously assigned explore/exploit states for each trial, we examined the action
112 associated with each choice, taking advantage of logging the coordinate locations of nosepokes
113 in our touchscreen operant chambers. This allowed us to have a two dimensional location for
114 each decision a mouse made across the entire touchscreen space. We started with an
115 Euclidean analysis to quantify the distance between successive touch responses where T1 was
116 compared to T2, T2 was compared to T3, T3 was compared to T4, so long as all touches were
117 from the same choice aperture and state (**Figure 1d**; (Ebitz and Hayden, 2021; Walther et al.,
118 2016)). One mouse was excluded from Euclidean analyses as they never had a sequence of
119 choices on the same side in the same state consecutively. Distance between successive exploit
120 touches was smaller and therefore less variable than successive explore touches (**Figure 1e**,
121 GLM, main effect of state, $p < 0.001$). However, sex also played a role - female mice had
122 shorter distances between successive touches than male mice (**Figure 1e**, GLM, main effect of
123 sex, $p = 0.01$). These results suggest that exploit touches are more stereotyped and perhaps
124 represent a more automated behavioral response than the same choice made during
125 exploration, and suggests that these behaviors are more stereotyped overall in females than in
126 males.

127
128 Although these data suggest that exploit choices are more stereotyped than exploration,
129 Euclidean analysis can only compare distances between touches that are consecutively
130 occurring on the same side, and in the same explore/exploit state. An alternative approach for
131 calculating distance that permits all touches to remain in analysis is the Mahalanobis distance, a
132 method for finding the distance between a point and the center of a distribution (**Figure 1f**)
133 (Ebitz and Hayden, 2021; Walther et al., 2016). With Mahalanobis distance the entire cluster of
134 data points was analyzed for each choice aperture, including both explore and exploit touches.
135 We separated the population of touch responses into those happening in explore states and
136 those in exploit states, and calculated separate Mahalanobis distances for exploit and explore
137 touches from centroids within each left/right choice aperture, combining the data from both
138 apertures across all trials and sessions and getting an average distance for each animal. The
139 Mahalanobis distance of an average exploit touch from the centroid of all exploit touches was
140 smaller and less variable than the distance of an average explore touch from the explore
141 centroid (**Figure 1g**, GLM, main effect of state, $p < 0.001$). Unlike Euclidean analysis, we do not
142 find significant sex differences in Mahalanobis distances (sex was dropped in the GLM model
143 with the lowest AIC value). The difference between sex influences on Euclidean and
144 Mahalanobis distances may reflect the trial-to-trial variability that Euclidean analysis captures
145 versus the overall distribution captured by Mahalanobis analysis. However, both analyses reveal
146 a main effect of explore/exploit state on touch variability - that exploit touches occur closer
147 together in space with less variability than explore touches.

148
149 In maze tasks, as animals approach a choice point, they exhibit a behavior called vicarious trial
150 and error (VTE) in which they move their head while surveying options to guide flexible decision
151 making, that is reduced as choices become repetitive (George et al., 2023; Johnson and
152 Redish, 2007; Redish, 2016; Tolman, 1948, 1939). This raised the possibility that in a
153 touchscreen environment, flexible decision making may be reflected in the approach to the
154 screen, allowing them to survey choices from a central location while exploring versus
155 approaching directly towards one option when exploiting. To determine whether our mice might
156 be exhibiting physical signs of deliberation between the left and right choice apertures during
157 the explore state, we calculated the distance from the midpoint of the entire touchscreen
158 between the two response apertures (**Figure 1h**). Explore touches happen significantly closer to
159 the center of the screen, and thus closer to the opposite response aperture, than exploit touches
160 (**Figure 1i**, GLM, main effect of state, $p < 0.001$). This did not differ by sex (GLM, no main effect
161 of sex, $p = 0.767$). These results suggest that in an explore state mice exhibit a VTE-like
162 behavior as they approach an area equidistant from both response apertures and deliberate
163 between left and right choice. Conversely, in an exploit state, mice make responses committed
164 to one aperture at a farther distance from the center of the screen.

165 **Previous reward is associated with reduced action variability separate from the effect of** 166 **explore/exploit state**

167 One potentially significant difference between explore and exploit states that might influence
168 animal actions is a differing rate of reward across states. Exploit behavior is likely to result from
169 prior success in obtaining reward, and thus exploit states might be expected to be associated

170 with higher reward. Alternatively, reward may have a separate impact on action precision that is
171 unrelated to explore/exploit state influences (Abe et al., 2011; Cashaback et al., 2017; Galea et
172 al., 2015; Hasson et al., 2015; Izawa and Shadmehr, 2011; Nikooyan and Ahmed, 2015;
173 Ramkumar et al., 2016; Therrien et al., 2016; Trommershäuser et al., 2003). To examine the
174 impact of reward on touch location we separated trials by outcome: rewarded/not rewarded. To
175 determine the impact of being rewarded on a previous trial, distance measurements were taken
176 between one trial back (T_{-1}) - labeled as *rewarded* or *non-rewarded* - and the current trial (T_0).
177 Euclidean and Mahalanobis distances for touches on trials following rewarded choices was
178 smaller and less variable than those following non-rewarded touches (**Figure 2a**, GLM, main
179 effect of reward, $p < 0.001$; **Figure 2c**, GLM, main effect of reward, $p < 0.001$). However, the
180 effect of reward on action precision was independent of an effect of explore/exploit state on
181 action precision, with both previous trial reward and explore/exploit state contributing main
182 effects on the precision of choice responses (**Figure 2b**, GLM, main effect of reward, $p < 0.001$;
183 **Figure 2d**, GLM, main effect of state, $p < 0.001$). Euclidean effects were stronger in females
184 (**Figure 2b**, GLM, main effect of sex, $p = 0.01$ and a sex by state interaction **Figure 2b**, GLM,
185 sex/state interaction, $p = 0.039$). As expected from prior Mahalanobis analysis, there was no
186 influence of sex on Mahalanobis distances. These results suggest that while reward impacts
187 touch location and minute adjustments in responding on the touchscreen, it does not overpower
188 the state effects shown in **Figure 1**.

189
190 In addition to modeling decision making behavior via Hidden Markov Models, we previously
191 used reinforcement learning models to assess sex differences in latent parameters that could
192 influence choice behavior, including learning rate parameter (α). We previously found in the
193 animals in the current dataset that the α parameter was significantly higher in females,
194 suggesting greater trial-to-trial influences of outcome on a female mouse's next choice than on
195 a male's. Euclidean distance between touches is reduced by female sex, reward, and exploit
196 behavior, and is a measure of trial-to-trial action variability. Therefore, we asked whether trial-to-
197 trial action variability as measured by Euclidean distance between sequential touches on either
198 aperture was correlated with trial-to-trial outcome sensitivity as measured by the α
199 parameter for the best fit reinforcement learning model from (Chen et al., 2021b). With sex,
200 distance, and α parameters as fixed effects, and individual mouse as a random effect, the
201 GLM revealed a higher α parameter, indicating more rapid outcome sensitivity/value
202 updating/learning rate, was associated with smaller distances between successive touches
203 (GLM, main effect of α , $p = 0.046$), suggesting that animals that were more sensitive to
204 outcomes in their choice behavior as measured by a reinforcement learning model also show
205 greater precision of their actions. Additionally, we replicated the sex difference in touch
206 precision with females having shorter distances (GLM, main effect of sex, $p = 0.018$).

207 **Exploit states and female sex reduce centroid shifting across session**

208 Given the difference between sex influences and the consistency of state influences on
209 Euclidean (**Figure 1e**) and Mahalanobis (**Figure 1g**) distances, we wanted to determine if the
210 pattern of responding shifts differently across a session for male/female mice and
211 explore/exploit state. Given that during exploration, animals are more likely to make choices
212 closer to the midpoint of the screen (**Figure 1j**), it could be the case that exploration can be

213 seen in terms of not only which aperture is chosen, but what section of the aperture responses
214 in an explore state center on. Although we observe animals sampling between two response
215 locations (right and left) in our decision task, it is not clear whether animals are behaving as
216 though they are sampling two discrete options versus sampling touching an area in space. To
217 measure this we separated each session into state *bouts*. A *bout* is defined as a period of
218 touches within one state on a particular choice aperture. State transition trials from either
219 explore to exploit or exploit to explore trigger a new *bout*. To determine how our mice use the
220 available space within the choice aperture we calculated area and perimeter associated with
221 each state bout. Bouts of touches were plotted and overlaid onto 2D contour plots from Plotly
222 Graphing Libraries (**Figure 3a**). For each bout, Open Source Computer Vision (OpenCV) was
223 used to capture the contours (bin traces) along continuous boundaries of the contour plots and
224 calculate area and perimeter for the outermost bin - which is recognized as the outer range of
225 nosepoke responses.

226
227 Regarding the area of the touchscreen choice apertures used by the mice, exploit bouts
228 occupied a smaller area (mm^2) on the screen and were less variable than explore bouts (**Figure**
229 **3b**, GLM, main effect of state, $p = 0.006$). Female mice used significantly less area of the
230 screen per bout than males (**Figure 3b**, GLM, main effect of sex, $p < 0.001$). The model used
231 included an interaction term between state and sex, which was not significant (**Figure 3b**, GLM,
232 interaction state/sex, $p = 0.989$). Perimeter of the touchscreen choice apertures used by the
233 mice, exploit bouts occupied a smaller boundary (mm) on the screen and were less variable
234 than explore bouts (GLM, main effect of state, $p < 0.001$). Female mice occupied a smaller
235 boundary on the screen and were less variable than bouts by male mice (GLM, main effect of
236 sex, $p = 0.004$). The model used included an interaction term between state and sex, which was
237 not significant (GLM, interaction state/sex, $p = 0.168$). Further suggesting differences in
238 touchscreen navigation across state and sex, where exploration and males interact with more
239 overall area of the screen.

240
241 Each new bout of responding includes its own centroid, and these centroids may minutely move
242 across the screen throughout a session, adjusting based on past experience. In a combination
243 of analysis techniques, **Figure 3c** shows how the distance between successive centroids is
244 calculated using the x,y centroid coordinates - as determined by the Mahalanobis analysis.
245 Distances between centroids for successive exploit bouts were smaller and less variable than
246 distances between centroids for successive explore bouts (**Figure 3d**, GLM, main effect of
247 state, $p < 0.001$). We found that touches occurring during one bout of exploration were farther
248 and more variable in distance from other bouts of exploration compared to more similar touch
249 patterns across bouts of exploitation. Given that mice are using more overall screen space
250 during explore than exploit trials, this further increases the likelihood that mice may be exploring
251 individual touch locations over and above sampling just the left/right options we define.

252 Discussion

253 The explore/exploit tradeoff is a fundamental property of choice selection during reward-guided
254 decision making. Explore and exploit states are mediated by distinct neural circuit activity and

Touchscreen response precision is sensitive to the explore/exploit tradeoff 8

255 reflect slower versus faster decision processes (Ebitz et al., 2020, 2019, 2018; Wang et al.,
256 2023; Wyatt et al., 2023) and so likely reflect different levels of subjective certainty in a choice.
257 Here, we take advantage of the observation that higher certainty actions in perceptual tasks are
258 more precise to ask whether exploit states, reward feedback, or other factors lead to increased
259 precision of choices. Using touchscreen operant chambers in mice we asked whether
260 explore/exploit balance governed the precision of actions during decision making, finding
261 independent effects of (1) explore/exploit state, (2) prior reward, and (3) sex on increasing
262 similarity of touches. These data suggest multiple independent mechanisms regulate the
263 precision of actions associated with choices and that the explore/exploit state is visible at the
264 level of motor performance.

265
266 Perceptual decision making tasks reveal that higher certainty decisions are more motorically
267 precise, even when the decision does not require motor accuracy (Follman et al., 2023; Palser
268 et al., 2018; Sanchez et al., 2024; Wolpert and Landy, 2012). One striking result from our study
269 is that exploit states reduce action variability in choice location. This suggests that exploit
270 touches reflect higher certainty in the animal. The similarity of touch locations suggests that
271 exploit decisions are more repetitive, stereotyped, or automated behavioral responses (Dezfouli
272 and Balleine, 2012; Dolan and Dayan, 2013; Gillan et al., 2016; Yin et al., 2004). Exploit choices
273 happen faster in comparison to explore choices (Chen et al., 2023b, 2021c; Ebitz et al., 2018),
274 an expression of these cognitive strategies at the motor level (Carsten et al., 2023; Chen et al.,
275 2017). Stereotyped performance of a behavior has previously been linked to a lack of
276 deliberation (Foster, 1998; Graybiel, 2008; Mitchell and Etches, 1977; Smith and Graybiel,
277 2016). Our findings are broadly consistent with the idea that exploit choices reflect behavioral
278 automation of a higher confidence response, while explore reflects deliberation.

279
280 Exploration and deliberation processes involve the subject surveying options (Gilbert and
281 Wilson, 2007; Payne et al., 1993; Rangel et al., 2008). Deliberation is physically expressed
282 through pausing, slower decision making, and “vicarious trial and error” behavior, reflecting
283 forward thinking and prospective deliberation (Dolan and Dayan, 2013; George et al., 2023;
284 Johnson and Redish, 2007; Redish, 2016; Tolman, 1948, 1939). We observed that explore
285 touches happen significantly closer to the center of the screen than exploit touches, which
286 implies animals are approaching exploratory choices between the two apertures, rather than
287 from off to one side. In addition, we found that touches occurring during one “bout” of
288 exploration were farther from other bouts of exploration compared to exploit. Given that mice
289 are using more overall screen space during explore than exploit trials, this suggests mice may
290 be exploring individual touch locations across the screen over and above sampling just the
291 left/right options we define. Self-directed exploration may reflect an increasingly fine-grained
292 goal-directed search for the most rewarding action, similar to autoshaping.

293
294 A potential confound between explore/exploit state and action precision is that exploit actions
295 are more likely to be reinforced. However, exploit states and prior reward independently
296 reduced action variability. This suggests that while reward may cause trial-to-trial adjustments in
297 responding on the touchscreen, reward does not overpower the state effect. Reward-triggered
298 changes in response precision may be a function of individual reward sensitivity. Animals with a

299 higher learning rate derived from a reinforcement learning model showed smaller distances
300 between successive touches, suggesting that reward sensitivity varying across individuals is
301 associated with increased action precision. This effect was larger in females than in males,
302 highlighting sex as a third independent factor governing choice precision.

303
304 The data in this manuscript were previously used to reveal a sex difference in the balance of
305 explore/exploit strategies (Chen et al., 2021c). Because male and female mice employ different
306 strategies in the two-arm restless bandit task, we sought to test whether motor responses
307 associated with the different strategies were physically different in distribution and spatial
308 location. We found that actions were more precise in females compared to males, independent
309 of the impact of explore/exploit state and reward experience, suggesting individual differences
310 regulating action precision over and above moment to moment features of the task. However,
311 not all explore/exploit differences were sex different. In particular, there was no sex difference in
312 how close animal responses were to the center of the screen during exploration. This suggests
313 that the overall deliberative process of an exploratory decision is probably similar across sexes,
314 but the sequential execution of these decisions are more similar in females than males. Overall
315 these findings agree with a growing literature that finds male decision and/or motor behavior to
316 be more variable than females in rodents (Chen et al., 2021a; Levy et al., 2023) and humans
317 (Dosenbach et al., 2017).

318
319 Touchscreens are increasingly used not only by rodent researchers, but by people working with
320 humans via smartphone-mediated ecological assessments. Our analysis reveals a powerful way
321 to evaluate the distribution and consistency of motor behaviors in choice responding. Motor
322 abnormalities are a common feature across patients with psychosis (Walther and Mittal, 2017),
323 autism (Mody et al., 2017; Mosconi and Sweeney, 2015), and depression (Sobin and Sackeim,
324 1997), and explore/exploit tradeoffs reveal neuropsychiatric influences (Addicott et al., 2017;
325 Wyatt et al., 2023). The increasing prevalence of touchscreen phone testing in human
326 neuropsychiatric research raises the distinct possibility of analyses of touch responses (Azenkot
327 and Zhai, 2012; Gosling and Mason, 2015; Harari et al., 2016; Intarasirisawat et al., 2019;
328 Miller, 2012) as a novel cross-species translational measure of explore/exploit tradeoffs.
329

330 **Methods**

331 **Subjects**

332 Animals were thirty-two 129/B6J F1 mice (16 male and 16 female) from The Jackson
333 Laboratory. Behavioral data from these mice running this task were previously published by the
334 lab (Chen et al., 2021b). Colony rooms were temperature controlled (20.5°C; 69°F) and on a
335 light-dark cycle of 12 hours with the lights off at 9am. Mice were housed in groups of four with
336 water ad libitum. Mice were food restricted to no lower than 85% of their free-feeding body
337 weight. All animals were cared for according to the guidelines of the National Institution of
338 Health and the University of Minnesota (UMN) and UMN IACUC approval.

339 Behavioral Data

340 Details on the methods for behavioral training and the restless bandit task are published in
341 (Chen et al., 2021b). Behavioral testing was carried out in the same touchscreen chambers for
342 all mice throughout the present study (Lafayette Instrument Company, Lafayette, IN).
343 Computational models were fit to mouse data in this paper, including a hidden Markov model
344 (HMM) and an RLCK reinforcement learning model (Chen et al., 2021b). The HMM was used to
345 determine when animals were exploring or exploiting their options in the restless bandit task,
346 where $P(\text{exploration})$ is the probability of mouse exploration between choices. The previous
347 manuscript compared several different RL models and identified the strongest fit to animal
348 behavior from an RLCK model, which captures both value-based and value-independent
349 decisions using the following four parameters: learning rate, decision noise, choice bias, and
350 choice stickiness. Here we use this RLCK model's alpha parameter compared to distance
351 between successive touches to assess how learning rate impacts micro adjustments to spatial
352 touch locations across sex. For validation of both models please see (Chen et al., 2021b) eLife
353 publication.

354 Coordinate Analysis

355 The Bussey-Saksida touchscreen apparatus (Lafayette Instrument Company) is sensitive to
356 continuous and rapidly repeated touches in the same location and across the entirety of the
357 screen (Heath et al., 2015). Each touchscreen represents the x,y coordinates of each response
358 an animal makes on the screen from IR beam technology where IR emitters are positioned
359 along two sides of the screen (i.e. top and right sides) and IR receivers are positioned along the
360 other two sides of the screen (i.e. bottom and left sides). In this configuration, IR beams are
361 ideally suited to determine the shadow of the touch to triangulate the location of choice
362 response. IR beam configuration results in a touch resolution that matches the monitor
363 resolution of 800x600 pixels. **Figure 1b** visualizes this data, representing the choices of four
364 different mice selecting between two options on the touchscreen over 300 trials, with explore
365 responses in the lighter purple and exploit responses in the darker purple. **Figure 1b** provides
366 an example of nosepoke responses for one mouse across a session and the change in touch
367 pattern between explore/exploit touches as identified by our HMM. Left and right touchscreen
368 choice apertures are 240x240 pixels each, never change position or size, and x,y coordinates
369 are separately generated for each touch aperture. Throughout all analyses we have transformed
370 pixels into millimeters. 1 pixel is 0.29 millimeters. Unless mentioned otherwise, for all data, a
371 GLM stepwise model selection analysis was used to determine the optimal model with the
372 lowest AIC value and p values are shared from those most optimal models.

373 Distance from the Center of the Screen

374 The spatial split in exploration and exploitation visualized by these plots suggested that explore
375 trials were closer to the center of the touchscreen than exploit trials were, prompting us to
376 quantify the distances (**Figure 1i**). With the center of the screen being 400 out of 800 total pixels
377 (width of the screen), the difference between the x pixel coordinate of the x,y location of each
378 touch response and 400 pixels was calculated and converted into millimeters. An absolute value
379 is applied so that the distance away from the center of the screen is always a positive value to

380 reflect distance. This calculation was done across all touches in every session. Trials were split
381 by explore and exploit and all data was averaged across all eight restless bandit sessions for
382 graphing purposes.

383

384 Example (x,y) is (34,208).

385 Distance from the center of the screen = $|400 - x|$

386 Distance from the center of the screen = $|400 - 34| = 366$ pixels.

387 **Euclidean Analysis**

388 The first method we used to quantify the distance between nosepoke touches was a Euclidean
389 Analysis (Ebitz and Hayden, 2021; Walther et al., 2016) in which we used the pythagorean
390 theorem to calculate the hypotenuse between two points with (x,y) coordinates that were
391 successive, from the same choice aperture (left/right), and within the same HMM decision state
392 (explore/exploit) (**Figure 1d**). In python this calculation was done using `numpy.hypot()`. A
393 drawback of this analysis is the amount of data points that get excluded given that the included
394 data points must be consecutively from the same choice aperture side and within the same
395 state. Distances were split by explore and exploit and all data was averaged across all eight
396 restless bandit sessions for graphing purposes. In the example below “T” represents touch
397 (nosepoke).

398

399 Example T_1 is (x_1, y_1) and T_2 is (x_2, y_2) .

400 Distance Between Successive Touches (hypotenuse) = $\sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2)}$

401 **Mahalanobis Analysis**

402 The second method we used to quantify touch patterns was a Mahalanobis analysis (Ebitz and
403 Hayden, 2021; Walther et al., 2016) where, unlike the Euclidean analysis, we didn't have to
404 exclude any touch data points. With this analysis we were able to calculate separate centroids
405 based on the data clusters for both the left side touches and right side touches and calculate the
406 distance of each touch coordinate from each overall centroid (**Figure 1f**). The centroid is the
407 central point in the data field that can be considered the overall mean for multivariate data given
408 that this is the point where all means from all variables intersect. The further away a data point
409 (touch) is from the centroid, the larger the Mahalanobis distance value. Distances were split by
410 explore and exploit and all data was averaged across all eight restless bandit sessions for
411 graphing purposes. In the formula below X_A and X_B represent a pair of objects, which are the x
412 and y coordinates; C is the sample covariance matrix, calculated using `numpy.cov()` in python;
413 and T is the transposition of the matrix over its diagonal, calculated using `numpy.linalg.inv()` in
414 python.

415

416 Mahalanobis Distance = $[(X_B - X_A)^T * C^{-1} * (X_B - X_A)]^{0.5}$

417 **Reward**

418 To determine whether being rewarded in the restless bandit task impacts touch location, we
419 compared trial outcome (rewarded or non-rewarded) from the previous trial (T_{-1}) to the change

420 in touch location on the current trial (T_0). This was done using both Euclidean and Mahalanobis
421 analyses.

422 **Distance Between Successive Bouts**

423 To understand how touches were organized within and across periods of exploration or exploit
424 as defined by HMM, we divided the data into “bouts”. Rather than looking at our nosepoke data
425 clusters throughout an entire session, a “bout” is described as a period of touches within one
426 HMM defined behavioral state on one particular choice aperture. Thus, explore states may
427 contain separate bouts on the left or right side, but these are analyzed separately. State
428 transition trials from either explore to exploit or exploit to explore trigger a new “bout.” By looking
429 at individual state bouts of choice responding, we can investigate whether explore or exploit
430 centroids on a given response area are shifting more throughout a session. This analysis
431 combines both Euclidean and Mahalanobis methods previously described. Mahalanobis
432 analysis is used to determine the centroid of each individual “bout.” From here, the distance
433 between successive centroids is calculated using the Euclidean analysis, which employs the
434 pythagorean theorem (**Figure 3c**). Distances were split by explore and exploit and all data was
435 averaged across all eight restless bandit sessions for graphing purposes. In the example below
436 “C” represents centroid.

437
438 Example C_1 is (x_1, y_1) and C_2 is (x_2, y_2) .

439 Distance Between Successive Touches (hypotenuse) = $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

440 **Contour Plots and Area Calculations**

441 In order to calculate the amount of space occupied by each bout we calculated the area and
442 perimeter of the bouts. In Python, 2D contour plots from Plotly Graphing Libraries were fit over
443 our nosepoke touch locations to visualize the density and range of choice responding. Bins
444 edges were designated by numpy.histogram and filtered at every-other bin so they were twice
445 as big as the standard output. The color bar was fixed from 0 to 1 across all generated plots to
446 ensure consistency of calculations (**Figure 3a**). Contour fill was removed, leaving just the
447 outlines at a thickness of “3” so the trace would be better recognized by OpenCV.

448
449 Once a contour plot was generated for each bout, Open Source Computer Vision (OpenCV)
450 was used to capture the contours along continuous boundaries and calculate area
451 (`cv.contourArea`) and perimeter (`cv.arcLength`) for each bin. While tracing the contours,
452 `cv.threshold` was set to `cv.THRESH_BINARY` and `cv.findContours` was set to
453 `cv.CHAIN_APPROX_SIMPLE`. Contour Approximation was used when it was necessary to
454 approximate the area between two separate contour groups. We focused on the dimensions of
455 the outermost bin as the best representation for the spread of data throughout a bout (**Figure**
456 **3a**). The outermost bin was filtered using the structure hierarchy, or rather the nested orientation
457 of the contours labeled numerically with “parent” and “child” identifications. Areas and
458 perimeters of bouts were split by explore and exploit and all data was averaged across all eight
459 restless bandit sessions for graphing purposes. Finally, area and perimeter were calculated for
460 the correctly identified contour bin. OpenCV was run through Minnesota Supercomputing
461 Institute (MSI).

462 Acknowledgments

463 This work was supported by NIMH R01 MH123661, NIMH P50 MH119569, Canada Research
464 Chair CRC-2022-00192 (RBE), NSERC RGPIN-2020-05577 (RBE), NIMH T32 training grant
465 MH115886, NIDA T32 training grant DA050560, and NIDA T32 training grant DA007234. We
466 thank Nic Glewwe for comments on an earlier draft of the manuscript and Matt Croxall from
467 Lafayette Instrument Company for valuable feedback and technical support.

468 References

- 469 Abe M, Schambra H, Wassermann EM, Luckenbaugh D, Schweighofer N, Cohen LG. 2011.
470 Reward improves long-term retention of a motor memory through induction of offline
471 memory gains. *Curr Biol* **21**:557–562.
- 472 Addicott MA, Pearson JM, Sweitzer MM, Barack DL, Platt ML. 2017. A Primer on Foraging and
473 the Explore/Exploit Trade-Off for Psychiatry Research. *Neuropsychopharmacology*
474 **42**:1931–1939.
- 475 Alabi OO, Fortunato MP, Fuccillo MV. 2019. Behavioral Paradigms to Probe Individual Mouse
476 Differences in Value-Based Decision Making. *Front Neurosci* **13**:50.
- 477 Azenkot S, Zhai S. 2012. Touch behavior with different postures on soft smartphone
478 keyboards Proceedings of the 14th International Conference on Human-Computer
479 Interaction with Mobile Devices and Services. Presented at the MobileHCI '12: 14th
480 International Conference on Human Computer Interaction with Mobile Devices and
481 Services. New York, NY, USA: ACM. doi:10.1145/2371574.2371612
- 482 Badre D, Doll BB, Long NM, Frank MJ. 2012. Rostrolateral prefrontal cortex and individual
483 differences in uncertainty-driven exploration. *Neuron* **73**:595–607.
- 484 Bari BA, Grossman CD, Lubin EE, Rajagopalan AE, Cressy JI, Cohen JY. 2019. Stable
485 Representations of Decision Variables for Flexible Behavior. *Neuron* **103**:922–933.e7.
- 486 Capshew JH. 1993. Engineering behavior: Project Pigeon, World War II, and the conditioning of
487 B. F. Skinner. *Technol Cult* **34**:835–857.
- 488 Carsten T, Fievez F, Duque J. 2023. Movement characteristics impact decision-making and vice
489 versa. *Sci Rep* **13**:3281.
- 490 Cashaback JGA, McGregor HR, Mohatarem A, Gribble PL. 2017. Dissociating error-based and
491 reinforcement-based loss functions during sensorimotor learning. *PLoS Comput Biol*
492 **13**:e1005623.
- 493 Cavanagh JF, Figueroa CM, Cohen MX, Frank MJ. 2012. Frontal theta reflects uncertainty and
494 unexpectedness during exploration and exploitation. *Cereb Cortex* **22**:2575–2586.
- 495 Chen CS, Ebitz RB, Bindas SR, Redish AD, Hayden BY, Grissom NM. 2021a. Divergent
496 Strategies for Learning in Males and Females. *Curr Biol* **31**:39–50.e4.
- 497 Chen CS, Knep E, Han A, Ebitz RB, Grissom NM. 2021b. Sex differences in learning from
498 exploration. *Elife* **10**. doi:10.7554/eLife.69748
- 499 Chen CS, Knep E, Han A, Ebitz RB, Grissom NM. 2021c. Sex differences in learning from
500 exploration. *Elife* **10**. doi:10.7554/eLife.69748
- 501 Chen CS, Mueller D, Knep E, Ebitz RB, Grissom NM. 2023a. Dopamine and norepinephrine
502 differentially mediate the exploration-exploitation tradeoff. *bioRxiv*.
503 doi:10.1101/2023.01.09.523322
- 504 Chen CS, Mueller D, Knep E, Ebitz RB, Grissom NM. 2023b. Dopamine and norepinephrine
505 differentially mediate the exploration-exploitation tradeoff. *bioRxiv.org*.
506 doi:10.1101/2023.01.09.523322
- 507 Chen X, Mohr K, Galea JM. 2017. Predicting explorative motor learning using decision-making

- 508 and motor noise. *PLoS Comput Biol* **13**:e1005503.
- 509 Cohen JY. 2015. Dopamine and serotonin signals for reward across time scales. *Science*
- 510 **350**:47.
- 511 Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and
- 512 dorsolateral striatal systems for behavioral control. *Nat Neurosci* **8**:1704–1711.
- 513 Dezfouli A, Balleine BW. 2012. Habits, action sequences and reinforcement learning. *Eur J*
- 514 *Neurosci* **35**:1036–1051.
- 515 Dolan RJ, Dayan P. 2013. Goals and habits in the brain. *Neuron* **80**:312–325.
- 516 Dosenbach NUF, Koller JM, Earl EA, Miranda-Dominguez O, Klein RL, Van AN, Snyder AZ,
- 517 Nagel BJ, Nigg JT, Nguyen AL, Wesevich V, Greene DJ, Fair DA. 2017. Real-time motion
- 518 analytics during brain MRI improve data quality and reduce costs. *Neuroimage* **161**:80–93.
- 519 Ebitz RB, Albarran E, Moore T. 2018. Exploration Disrupts Choice-Predictive Signals and Alters
- 520 Dynamics in Prefrontal Cortex. *Neuron* **97**:450–461.e9.
- 521 Ebitz RB, Hayden BY. 2021. The population doctrine in cognitive neuroscience. *Neuron*
- 522 **109**:3055–3068.
- 523 Ebitz RB, Sleezer BJ, Jedema HP, Bradberry CW, Hayden BY. 2019. Tonic exploration governs
- 524 both flexibility and lapses. *PLoS Comput Biol* **15**:e1007475.
- 525 Ebitz RB, Tu JC, Hayden BY. 2020. Rules warp feature encoding in decision-making circuits.
- 526 *PLoS Biol* **18**:e3000951.
- 527 Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge SY, Daw
- 528 ND, Tank DW, Witten IB. 2019. Specialized coding of sensory, motor and cognitive
- 529 variables in VTA dopamine neurons. *Nature* **570**:509–513.
- 530 Follman EG, Chevée M, Kim CJ, Johnson AR, Tat J, Leonard MZ, Calipari ES. 2023. Task
- 531 parameters influence operant response variability in mice. *Psychopharmacology* **240**:213–
- 532 225.
- 533 Foster LG. 1998. Nervous habits and stereotyped behaviors in preschool children. *J Am Acad*
- 534 *Child Adolesc Psychiatry* **37**:711–717.
- 535 Frank MJ, Fossella JA. 2011. Neurogenetics and pharmacology of learning, motivation, and
- 536 cognition. *Neuropsychopharmacology* **36**:133–152.
- 537 Galea JM, Mallia E, Rothwell J, Diedrichsen J. 2015. The dissociable effects of punishment and
- 538 reward on motor learning. *Nat Neurosci* **18**:597–602.
- 539 George AE, Stout JJ, Griffin AL. 2023. Pausing and reorienting behaviors enhance the
- 540 performance of a spatial working memory task. *Behav Brain Res* **446**:114410.
- 541 Gilbert DT, Wilson TD. 2007. Propection: experiencing the future. *Science* **317**:1351–1354.
- 542 Gillan CM, Robbins TW, Sahakian BJ, van den Heuvel OA, van Wingen G. 2016. The role of
- 543 habit in compulsivity. *Eur Neuropsychopharmacol* **26**:828–840.
- 544 González VV, Zhang Y, Ashikyan SA, Rickard A, Yassine I, Romero-Sosa JL, Blaisdell AP,
- 545 Izquierdo A. 2024. A special role for anterior cingulate cortex, but not orbitofrontal cortex or
- 546 basolateral amygdala, in choices involving information. *Cereb Cortex* **34**.
- 547 doi:10.1093/cercor/bhae135
- 548 Goodale MA. 1983. Visually guided pecking in the pigeon (*Columba livia*). *Brain Behav Evol*
- 549 **22**:22–41.
- 550 Gosling SD, Mason W. 2015. Internet research in psychology. *Annu Rev Psychol* **66**:877–902.
- 551 Graybiel AM. 2008. Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* **31**:359–387.
- 552 Grissom NM, Reyes TM. 2019. Correction: Let's call the whole thing off: evaluating gender and
- 553 sex differences in executive function. *Neuropsychopharmacology* **44**:1344.
- 554 Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. 2016. Using
- 555 Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical
- 556 Considerations, and Challenges. *Perspect Psychol Sci* **11**:838–854.
- 557 Hassall CD, Holland K, Krigolson OE. 2013. What do I do now? An electroencephalographic
- 558 investigation of the explore/exploit dilemma. *Neuroscience* **228**:361–370.

- 559 Hasson CJ, Manczurovsky J, Yen S-C. 2015. A reinforcement learning approach to gait training
560 improves retention. *Front Hum Neurosci* **9**:459.
- 561 Heath CJ, Bussey TJ, Saksida LM. 2015. Motivational assessment of mice using the
562 touchscreen operant testing system: effects of dopaminergic drugs. *Psychopharmacology*
563 **232**:4043–4057.
- 564 Intarasirisawat J, Ang CS, Efstratiou C, Dickens LWF, Page R. 2019. Exploring the touch and
565 motion features in game-based cognitive assessments. *Proc ACM Interact Mob Wearable*
566 *Ubiquitous Technol* **3**:1–25.
- 567 Izawa J, Shadmehr R. 2011. Learning from Sensory and Reward Prediction Errors during Motor
568 Adaptation. *PLoS Comput Biol* **7**:e1002012.
- 569 Jager R, Zeigler HP. 1991. Visual field organization and peck localization in the pigeon
570 (*Columba livia*). *Behav Brain Res* **45**:65–69.
- 571 Johnson A, Redish AD. 2007. Neural ensembles in CA3 transiently encode paths forward of the
572 animal at a decision point. *J Neurosci* **27**:12176–12189.
- 573 Laureiro-Martínez D, Brusoni S, Zollo M. 2010. The neuroscientific foundations of the
574 exploration–exploitation dilemma. *J Neurosci Psychol Econ* **3**:95–115.
- 575 Levy DR, Hunter N, Lin S, Robinson EM, Gillis W, Conlin EB, Anyoha R, Shansky RM, Datta
576 SR. 2023. Mouse spontaneous behavior reflects individual variation rather than estrous
577 state. *Curr Biol* **33**:1358–1364.e4.
- 578 Miller G. 2012. The Smartphone Psychology Manifesto. *Perspect Psychol Sci* **7**:221–237.
- 579 Mitchell R, Etches P. 1977. Rhythmic habit patterns (stereotypies). *Dev Med Child Neurol*
580 **19**:545–550.
- 581 Mody M, Shui AM, Nowinski LA, Golas SB, Ferrone C, O'Rourke JA, McDougale CJ. 2017.
582 Communication Deficits and the Motor System: Exploring Patterns of Associations in
583 Autism Spectrum Disorder (ASD). *J Autism Dev Disord* **47**:155–162.
- 584 Mohebi A, Pettibone JR, Hamid AA, Wong J-MT, Vinson LT, Patriarchi T, Tian L, Kennedy RT,
585 Berke JD. 2019. Publisher Correction: Dissociable dopamine dynamics for learning and
586 motivation. *Nature* **571**:E3.
- 587 Mosconi MW, Sweeney JA. 2015. Sensorimotor dysfunctions as primary features of autism
588 spectrum disorders. *Sci China Life Sci* **58**:1016–1023.
- 589 Nikooyan AA, Ahmed AA. 2015. Reward feedback accelerates motor learning. *J Neurophysiol*
590 **113**:633–646.
- 591 Palser ER, Fotopoulou A, Kilner JM. 2018. Altering movement parameters disrupts
592 metacognitive accuracy. *Conscious Cogn* **57**:33–40.
- 593 Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, Daw ND, Witten IB. 2016.
594 Reward and choice encoding in terminals of midbrain dopamine neurons depends on
595 striatal target. *Nat Neurosci* **19**:845–854.
- 596 Payne JW, Bettman JR, Johnson EJ. 1993. *The Adaptive Decision Maker*. Cambridge
597 University Press.
- 598 Peterson GB. 2004. A DAY OF GREAT ILLUMINATION: B. F. SKINNER'S DISCOVERY OF
599 SHAPING. *J Exp Anal Behav* **82**:317–328.
- 600 Ramkumar P, Dekleva B, Cooler S, Miller L, Kording K. 2016. Premotor and Motor Cortices
601 Encode Reward. *PLoS One* **11**:e0160851.
- 602 Rangel A, Camerer C, Montague PR. 2008. A framework for studying the neurobiology of value-
603 based decision making. *Nat Rev Neurosci* **9**:545–556.
- 604 Redish AD. 2016. Vicarious trial and error. *Nat Rev Neurosci* **17**:147–159.
- 605 Sanchez R, Courant A, Desantis A, Gajdos T. 2024. Making precise movements increases
606 confidence in perceptual decisions. *Cognition* **249**:105832.
- 607 Skinner BF. 1960. Pigeons in a pelican. *Am Psychol* **15**:28–37.
- 608 Smith KS, Graybiel AM. 2016. Habit formation. *Dialogues Clin Neurosci* **18**:33–43.
- 609 Sobin C, Sackeim HA. 1997. Psychomotor symptoms of depression. *Am J Psychiatry* **154**:4–17.

- 610 Spetch ML, Cheng K, Mondloch MV. 1992. Landmark use by pigeons in a touch-screen spatial
611 search task. *Anim Learn Behav* **20**:281–292.
- 612 Stephens DW. 2008. Decision ecology: foraging and the ecology of animal decision making.
613 *Cogn Affect Behav Neurosci* **8**:475–484.
- 614 Therrien AS, Wolpert DM, Bastian AJ. 2016. Effective reinforcement learning following
615 cerebellar damage requires a balance between exploration and motor noise. *Brain*
616 **139**:101–114.
- 617 Ting C-C, Salem-Garcia N, Palminteri S, Engelmann JB, Lebreton M. 2023. Neural and
618 computational underpinnings of biased confidence in human reinforcement learning. *Nat*
619 *Commun* **14**:6896.
- 620 Tolman EC. 1948. Cognitive maps in rats and men. *Psychol Rev* **55**:189–208.
- 621 Tolman EC. 1939. Prediction of vicarious trial and error by means of the schematic sowbug.
622 *Psychol Rev* **46**:318–336.
- 623 Trommershäuser J, Maloney LT, Landy MS. 2003. Statistical decision theory and trade-offs in
624 the control of motor response. *Spat Vis* **16**:255–275.
- 625 Trudel N, Scholl J, Klein-Flügge MC, Fouragnan E, Tankelevitch L, Wittmann MK, Rushworth
626 MFS. 2021. Polarity of uncertainty representation during exploration and exploitation in
627 ventromedial prefrontal cortex. *Nat Hum Behav* **5**:83–98.
- 628 Walker AR, Navarro DJ, Newell BR, Beesley T. 2022. Protection from uncertainty in the
629 exploration/exploitation trade-off. *J Exp Psychol Learn Mem Cogn* **48**:547–568.
- 630 Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, Diedrichsen J. 2016. Reliability of dissimilarity
631 measures for multi-voxel pattern analysis. *Neuroimage* **137**:188–200.
- 632 Walther S, Mittal VA. 2017. Motor System Pathology in Psychosis. *Curr Psychiatry Rep* **19**:97.
- 633 Wang S, Falcone R, Richmond B, Averbach BB. 2023. Attractor dynamics reflect decision
634 confidence in macaque prefrontal cortex. *Nat Neurosci*. doi:10.1038/s41593-023-01445-x
- 635 Wershba A, Pleskac T. 2010. Making Assessments While Taking Sequential Risks.
636 *Proceedings of the Annual Meeting of the Cognitive Science Society* **32**.
- 637 Wolpert DM, Landy MS. 2012. Motor control is decision-making. *Curr Opin Neurobiol* **22**:996–
638 1003.
- 639 Wyatt LE, Hewan PA, Hogeveen J, Spreng RN, Turner GR. 2023. Exploration versus
640 exploitation decisions in the human brain: A systematic review of functional neuroimaging
641 and neuropsychological studies. *Neuropsychologia* **192**:108740.
- 642 Yin HH, Knowlton BJ, Balleine BW. 2004. Lesions of dorsolateral striatum preserve outcome
643 expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci* **19**:181–189.

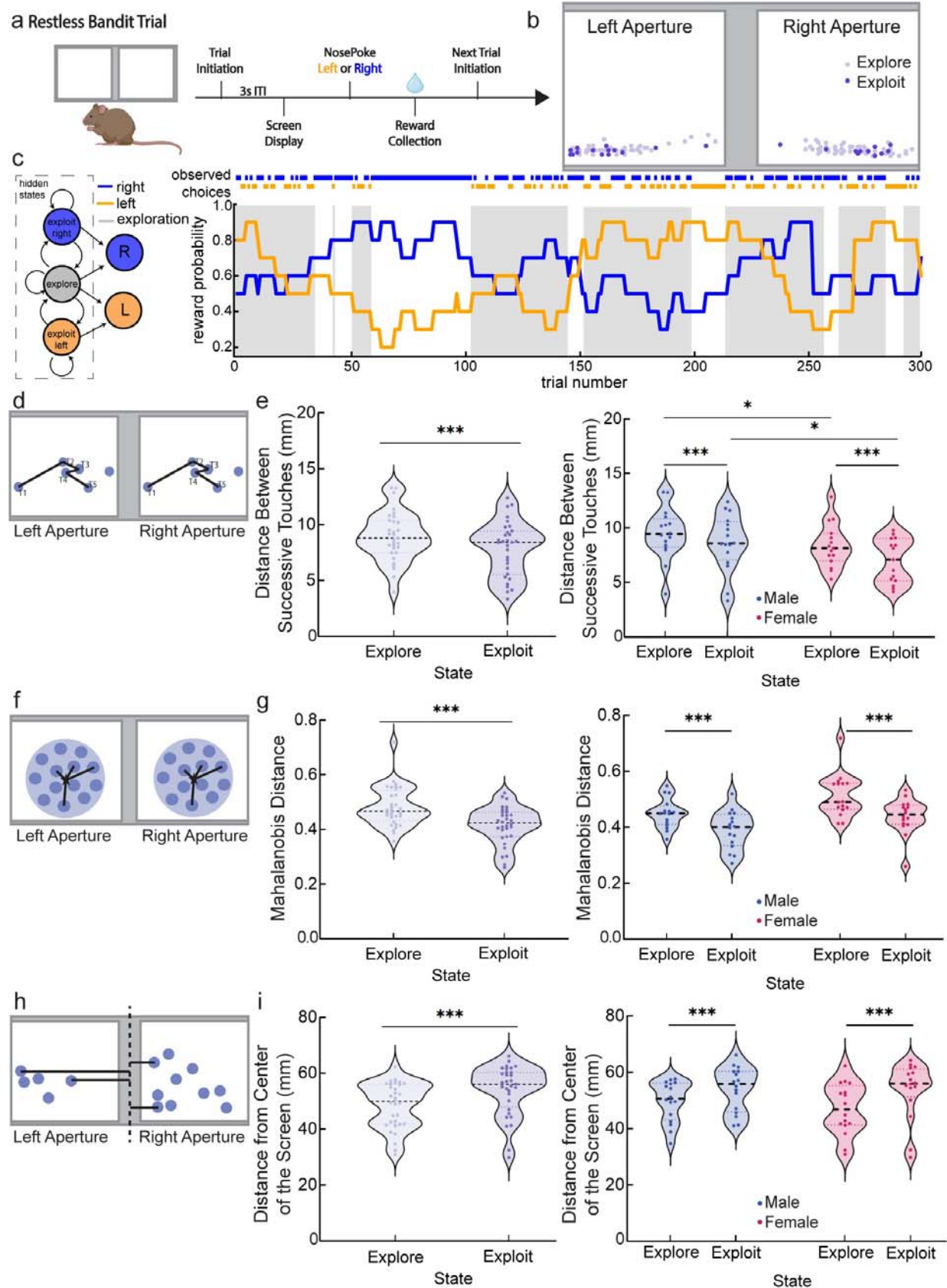
Touchscreen response precision is sensitive to the explore/exploit tradeoff

17

644 **Figures**

Touchscreen response precision is sensitive to the explore/exploit tradeoff

18

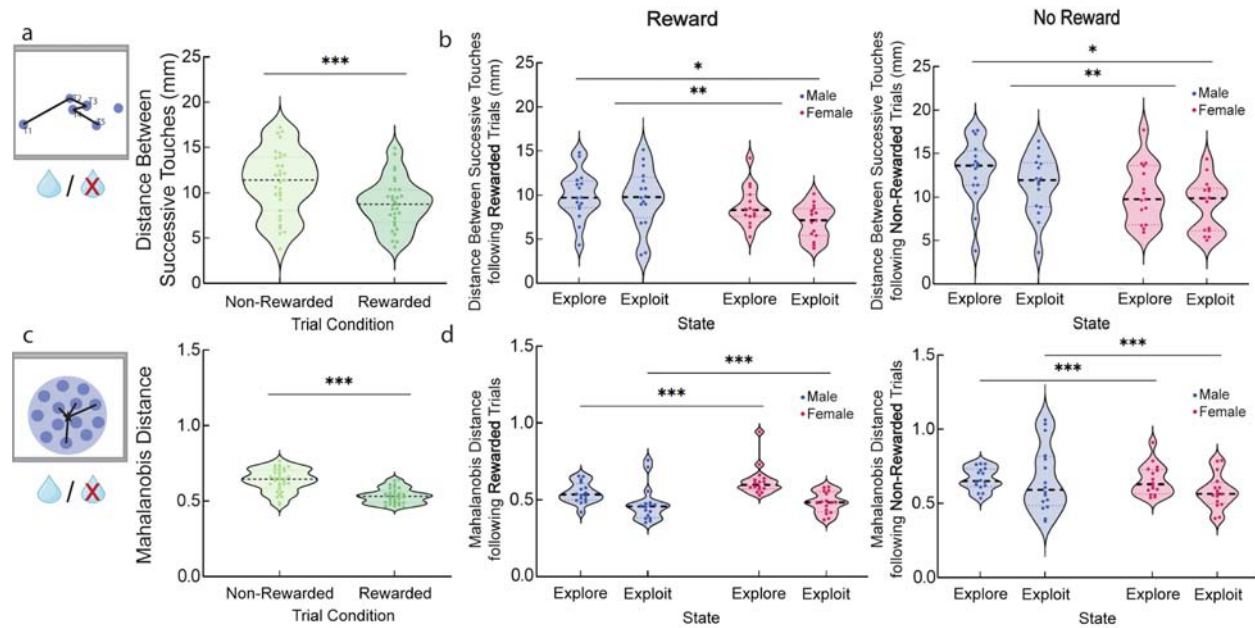


646 **Figure 1: Exploit states and female sex reduce action variability during decision making.**

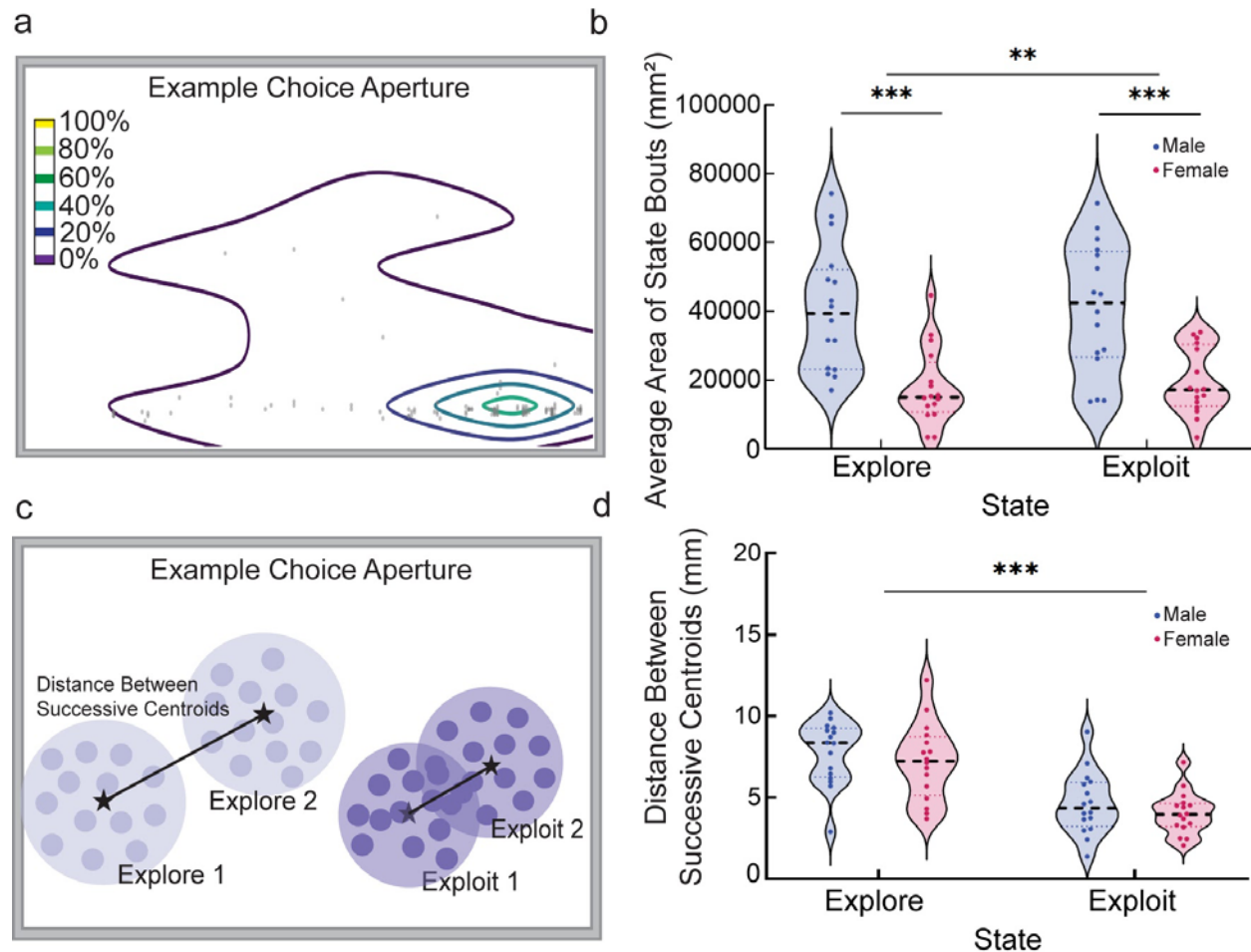
647 a) schematic depicting the timeline of a single trial. White squares indicate left/right spatial
648 choice. b) An example of touch screen responses from one animal and one session, where light
649 purple indicates explore touches and dark purple indicates exploit touches. c) Schematic
650 depicting the hidden Markov model (HMM) and labeling explore trials along an example two-arm
651 restless bandit probability walk. Orange traces indicate the probability and choices of left side
652 touches. Blue traces indicate the probability and choices of right side touches. Gray shaded
653 regions indicate HMM labeled explore trials. d) Schematic of Euclidean distance where the
654 distance is calculated between touch 1 and touch 2, touch 2 and touch 3, touch 3 and touch 4,
655 and so on. Shown here are possible left/right touches in blue and the distance relationship from
656 one to another represented by black lines. e) Average Euclidean distance split by state (left) and
657 sex (right). Exploit touches and females had significantly reduced Euclidean distance. Light
658 purple indicates distance between explore touches and dark purple indicates distance between
659 exploit touches. Red indicates female and blue indicates male mice. In violin graphs, individual
660 data points are data from one mouse averaged across all sessions. f) Schematic of
661 Mahalanobis distance where the individual data points are measured from the overall centroid of
662 the dataset. Shown here are possible left/right Mahalanobis clusters (light blue circles) and
663 centroids (stars) and the Mahalanobis distance relationship from each touch (darker blue
664 circles) in a cluster to the centroid represented by black lines. g) Average Mahalanobis distance
665 split by state (left) and sex (right). Exploit touches had significantly reduced Mahalanobis
666 distance. Light purple indicates Mahalanobis distance between explore touches and dark purple
667 indicates Mahalanobis distance between exploit touches. Red indicates female and blue
668 indicates male mice. h) Schematic of distance from the center of the screen where touch
669 distance from both left and right choice apertures is measured from the midpoint of the operant
670 screen. Shown here are possible left/right touches in blue and the distance of each from the
671 center of the touchscreen represented by black lines. i) Average distance from the center of the
672 screen split by state (left) and sex (right). Explore touches were significantly closer to the center
673 of the screen. Light purple indicates distance from the center of the screen for explore touches
674 and dark purple indicates distance from the center of the screen for exploit touches. Red
675 indicates female and blue indicates male mice. For simplicity of visualization, all plots are
676 averages across trials and sessions, so that each individual data point plotted represents the
677 overall average for a mouse. Significant throughout this paper is represented in the following
678 way: * p value less than 0.05 and greater than 0.01; ** p value less than 0.01 and greater than
679 or equal to 0.001; *** p value less than 0.001. Violin graphs depict median and quartiles of the
680 dataset.

Touchscreen response precision is sensitive to the explore/exploit tradeoff

20



681
682 **Figure 2: Previous reward reduces action variability independently from explore/exploit**
683 **balance or female sex.** a) Average Euclidean distance comparing rewarded vs. non-rewarded
684 trials. Touches following rewarded trials had significantly reduced Euclidean distance. Light
685 green indicates distance between non-rewarded touches and dark green indicates distance
686 between rewarded touches. In violin graphs, individual data points are data from one mouse
687 averaged across all sessions. b) Average Euclidean distance for rewarded (left) and non-
688 rewarded (right) trials split by state and sex. Exploit touches and females had significantly
689 reduced Euclidean distance. Red indicates female and blue indicates male mice. c) Average
690 Mahalanobis distance comparing rewarded vs. non-rewarded trials. Touches following rewarded
691 trials had significantly reduced Mahalanobis distance. Light green indicates Mahalanobis
692 distance between non-rewarded touches and dark green indicates Mahalanobis distance
693 between rewarded touches. d) Average Mahalanobis distance for rewarded (left) and non-
694 rewarded (right) trials split by state and sex. Exploit touches had significantly reduced
695 Mahalanobis distance. Red indicates female and blue indicates male mice. * p value less than
696 0.05 and greater than 0.01; ** p value less than 0.01 and greater than or equal to 0.001; *** p
697 value less than 0.001. Violin graphs depict median and quartiles of the dataset.



698

699

700 **Figure 3: Exploit states and female sex reduce the total response space chosen across a**

701 **decision making session.** a) An example 2D contour plot from Plotly Graphing Libraries fit

702 Small gray circles are nosepoke touches within the bout of response data. Color map

703 corresponds with density of data points within each bin, where the darkest purple (outer bin) is

704 the least dense contour bin, which is used to calculate area and perimeter of the bout. b)

705 Average area of bouts split by state and sex. Exploit touches and females had significantly

706 reduced area. Red indicates female and blue indicates male mice. In violin graphs, individual

707 data points are data from one mouse averaged across all sessions. c) Schematic depicting

708 centroid shifts, where the Euclidean distance between two successive Mahalanobis centroids is

709 calculated. Stars represent example centroids associated with bouts and black lines represent

710 the distance calculations between those centroids. d) Centroid shifts split by state and sex.

711 Centroid shifts were significantly smaller for exploit bouts. Red indicates female and blue

712 indicates male mice. * p value less than 0.05 and greater than 0.01; ** p value less than 0.01

713 and greater than or equal to 0.001; *** p value less than 0.001. Violin graphs depict median and

714 quartiles of the dataset.