

Many paths to the same goal: balancing exploration and exploitation during probabilistic route planning

<https://doi.org/10.1523/ENEURO.0536-19.2020>

Cite as: eNeuro 2020; 10.1523/ENEURO.0536-19.2020

Received: 18 December 2019

Revised: 21 April 2020

Accepted: 7 May 2020

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2020 Jackson et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Manuscript Title Page

1. Manuscript Title: **Many paths to the same goal: balancing exploration and exploitation during probabilistic route planning**

2. Abbreviated Title: **Explore vs. exploit during route planning**

3. List of all Author Name and Affiliations in order:

Brian J. Jackson¹, Gusti Lulu Fatima¹, Sujean Oh¹, David H. Gire^{1,*}

Affiliations:

¹Department of Psychology, University of Washington, Seattle, WA 98195, USA

4. Author contributions: **DHG and BJJ initiated the study and designed the experiments. DHG, BJJ and SO built the automated navigation arena and wrote the behavioral monitoring software. BJJ and GL performed experiments. DHG performed analysis of behavioral data with input from BJJ and GL. DHG and BJJ generated all figures reporting experimental data. DHG and BJJ wrote the text. All authors provided editing suggestions for the figures and text.**

5. Correspondence should be addressed to: **dhgire@uw.edu**

6. Number of Figures: **6**

7. Number of Tables: **0**

8. Number of Multimedia: **0**

9. Number of words for Abstract: **242**

10. Number of words for Significance Statement: **106**

11. Number of words for Introduction: **730**

12. Number of words for Discussion: **1,402**

13. Acknowledgements: **We thank Bingni Brunton, Venkatesh Gopal, and Agnese Seminara for helpful discussion and members of the Gire lab for comments on the manuscript. We would also like to thank Sarahi Carolina Ponton Junes and Ryan Van Ort for assistance with data collection.**

14. Conflict of Interest: **The authors report no conflict of interest.**

15. Funding Sources:

NIH/NIDCD grant R00 DC013305 (DHG)

NIH/NIDCD grant R21 DC018649 (DHG)

FACE Foundation (DHG)

University of Washington Innovation Award (DHG)

47 **Abstract: (242 words)**

48 During self-guided behaviors animals identify constraints of the problems they face and
49 adaptively employ appropriate strategies (Marsh, 2002). In the case of foraging, animals must
50 balance sensory-guided exploration of an environment with memory- guided exploitation of
51 known resource locations. Here we show that animals adaptively shift cognitive resources
52 between sensory and memory systems during foraging to optimize route planning under
53 uncertainty. We demonstrate this using a new, laboratory-based discovery method to define the
54 strategies used to solve a difficult route optimization scenario, the probabilistic “traveling
55 salesman” problem (Anaya Fuentes et al., 2018; Mukherjee et al., 2019; Raman & Gill, 2017).
56 Using this system, we precisely manipulated the strength of prior information as well as the
57 complexity of the problem. We find that rats are capable of efficiently solving this route-
58 planning problem, even under conditions with unreliable prior information and a large space of
59 possible solutions. Through analysis of animals’ trajectories, we show that they shift the
60 balance between exploiting known locations and searching for new locations of rewards based
61 upon the predictability of reward locations. When compared to a Bayesian search, we found
62 that animal performance is consistent with an approach that adaptively allocates cognitive
63 resources between sensory processing and memory, enhancing sensory acuity and reducing
64 memory load under conditions in which prior information is unreliable. Our findings establish
65 new approaches to understand neural substrates of natural behavior as well as the rational
66 development of biologically inspired approaches for complex real-world optimization.

67 **Significance Statement (106 words):**

68

69 Animals display remarkable problem-solving abilities across a variety of complex situations.
70 Here, we used a large, computer-controlled foraging field with precisely controlled probabilities
71 of food resources in either repeated or random locations to test how rats determine which
72 strategies to use to solve an extremely complicated route planning problem. We found that rats
73 balanced exploration for novel locations of food with exploitation of known food locations to
74 solve this problem, with the balance between exploratory and exploitative strategies governed
75 by the amount of information available regarding resource location. Our results show how
76 animals balance sensory input with learned information to solve complex, real-world route
77 planning problems.

78

79 **Introduction (730 words)**

80 Animals balance the ability to flexibly interact with their environment with the need to
81 reserve energy while foraging. Foraging in natural environments can be particularly difficult due
82 to the sparse and unreliable nature of sensory cues emanating from food sources. This is
83 especially true when animals need to travel between multiple locations and it is unknown
84 whether food will be present at these locations. Under conditions of high uncertainty it may be
85 beneficial to rely upon sensory information during foraging and utilize a more exploratory
86 approach, when the increased cognitive demand of this strategy is offset by the need to flexibly
87 interact with the environment. Conversely, using a memory-based strategy to exploit known
88 resource locations allows for the quick establishment of efficient stereotyped routes, yet result in
89 behaviors that are not readily adaptable to changing contingencies in the environment. It is
90 therefore important for animals to maintain cognitive flexibility while foraging in their natural
91 environment in order to execute the most efficient behaviors required for food procurement

92 (Dolan & Dayan, 2013). To this end, the ability to adaptively modify search strategy by using
93 internal representations of the dynamic environment would serve to vastly increase the
94 effectiveness of foraging bouts (Slotnick, 2001; S. Zhang & Manahan-Vaughan, 2015).

95 Animals must learn the constraints of their environment in order to determine how to
96 optimize their foraging strategies, with the balance of exploration versus exploitation being vital
97 in this context (Auh & Menguc, 2005; Gupta et al., 2006; Kramer & Weary, 1991; Mehlhorn et
98 al., 2015). During exploration, animals sample from multiple food patches over the course of
99 several foraging bouts. This allows them to construct an internal representation of different
100 possible locations where they can find food, with the benefit being that their future foraging
101 would be more resistant to reduced or noisy sensory cues. Exploitation of this information
102 follows and relies on remembering bountiful patch locations so that animals have a framework to
103 use for navigation. While benefits of exploitation include spending less energy traveling to
104 locations where it is unknown whether food will be available, potential drawbacks would be that
105 this strategy fails when resources have been exhausted or when resource locations change.
106 Additional exploration after establishing resource location is thus most useful when new resource
107 locations need to be discovered, such as when information regarding resource locations is found
108 to be unreliable. Under the constraints of foraging in an unpredictable environment, it is more
109 difficult to exploit reliable resource locations in order to reduce foraging costs and strategies
110 should shift toward exploration.

111 The ability to rapidly solve complex problems, such as optimization of foraging
112 strategies, is a defining feature of animal intelligence. Indeed, varieties of animals solve difficult
113 optimization problems nearly instantaneously (Drea & Carter, 2009; Kenward et al., 2005; Wall
114 & Balda, 1977; Q. Zhang et al., 2015). However, it has been difficult to study route optimization

115 during naturalistic foraging in a laboratory setting. Historically, many foraging tasks have been
 116 studied with apparatuses that do not explore the full behavioral repertoire of a natural forager.
 117 One issue is the difficulty of providing alternative possible paths for the animals when they are
 118 restricted to a track, such as a figure-8 maze (Pedigo et al., 2006). In these simplified tasks the
 119 space of available behaviors is limited to simple actions such as left and right turns. While other
 120 studies avoid these restrictions through the use of open field designs, these approaches
 121 necessarily reduce the precision and reproducibility of resource locations (Agarwal et al., 2014).
 122 We address these challenges by studying naturalistic foraging in a large, computer-controlled
 123 open field where food rewards can be precisely and reproducibly located anywhere in the
 124 environment.

125 Using our computer-controlled open field design we investigated the strategies rats use to
 126 solve a notoriously difficult optimization scenario, the probabilistic traveling salesman problem.
 127 In this problem, an agent must establish the most efficient (i.e., shortest) route between a finite
 128 number of locations and each location has a certain probability of containing pellets (Leipälä,
 129 1978; Percus & Martin, 1999). We observed rats' ability to follow efficient acquisition
 130 sequences and measured how well animal performance correlated with memory-guided
 131 exploitative strategies or sensory-guided exploratory strategies as a function of the predictability
 132 of the pellet distributions upon which animals were trained. These precise behavioral
 133 experiments suggest animals adaptively shift their reliance on sensory information in response to
 134 the reliability of the foraging environment.

135

136 **Materials and Methods**

137

138 **Subjects.** The experiments in this study were performed on 12 male Long-Evans rats, purchased

139 from Charles River Labs and housed individually. All animals were maintained on a 12-hour
140 reverse light-dark schedule (lights off at 7:00am) with *ad libitum* access to water. After a
141 weeklong habituation to the animal housing facility, all animals were then sustained at 85% of
142 their free- feeding body weight in order to maintain motivation. All tests were performed
143 between 9:00am and 6:00pm, during the dark phase of the light cycle. Zeitgeber Time (ZT, with
144 ZT0 = lights on in the animal facility) of experiments was ZT 14 to ZT 23. To limit distal visual
145 cues, all tests were performed under dim red light (~660 nm). All experimental procedures were
146 approved by the Institutional Animal Care and Use Committee at the University of Washington.

147
148
149 **Testing Apparatus.** The foraging arena was a large, fully enclosed open-field measuring 2.5m
150 in length, 1m in width, and 1m in height. The frame of the arena was constructed from T-slotted
151 aluminum railings. The sides of the arena were constructed from 1.27cm thick clear acrylic,
152 while the ceiling was 0.635cm in thickness. The floor was a sheet of 0.0635cm thick opaque
153 white acrylic. The ends of the arena were made from a wire mesh to allow for air to circulate
154 throughout. A nest area where the animals would remain during the intertrial interval was
155 attached to one end of the arena. The nest area was constructed from 1.27cm thick clear acrylic.
156 Two synchronized cameras (The Imaging Source; DMK 23UP1300; frame rate 120 per second)
157 were used to track the movement of the animals. An automated, custom-made pellet dispenser
158 was used to bait the arena with 45mg sucrose pellets (Bio-Serv). An Arduino Uno controlled the
159 movement of the motors running the pellet dispenser, allowing movement in the x- and y-
160 coordinate plane.

161 *Estimation of odor cues:* Odor cue dispersal in the arena was directly measured using an ethanol
162 source and miniature ethanol sensors (Tariq et al., 2019) that were scanned in a grid across the

163 arena. The maximal signal detected at each sensor location over 30 seconds was normalized and
164 reported in figure 4. There was no flow imposed on the arena, which limited the dispersal of
165 airborne odor cues.

166
167

168 **Behavioral Paradigm.** Before testing, all animals were habituated to the animal facility for 1
169 week. Animals then spent 2 days habituating to the attached waiting cage for ~15 minutes at a
170 time. In order to motivate animals to return to the waiting cage, sucrose pellets were placed in
171 the cage every 2 minutes when a 1 second, 1000Hz tone was played. They were then granted
172 access to the test arena and were given 2-3 days to habituate to it. Animals were considered to
173 have reached criterion when they were able to make 3 transitions between the waiting cage and
174 test arena within 30 minutes.

175 Animals were placed into the waiting cage at the beginning of each testing session. Rats
176 completed 1 session a day of 3 trials each. Before each trial, the automated pellet dispenser
177 baited the arena with sucrose pellets organized into 3 clusters of approximately 3 pellets each.
178 During foraging periods the dispenser was automatically lifted out of the arena so that the
179 animals could not interact with it. Procedures differed only through the testing phase, when
180 animals were assigned to forage within environments of high, medium, or low food location
181 predictability. Animals trained on the environment with high food location predictability (n=4)
182 were overtrained on a single distribution of pellet locations that stayed consistent across trials
183 and sessions. Animals foraging in the environment with low food location predictability (n=4)
184 were trained on unpredictable pellet distributions that changed across trials. All other animals
185 (n=4) were trained on a moderately predictable distribution of pellet locations that changed
186 slightly over time. All rats were given a maximum of 30 minutes to eat all of the sucrose pellets

187 during the session. The entire testing period lasted for 30-35 days with approximately 5 sessions
188 a week.

189 **Experimental Design and Statistical Analysis.** No explicit power analysis was conducted in
190 order to determine sample sizes. However, the number of animals used is consistent with
191 experiments in the current literature. All analyses were conducted using MATLAB (MathWorks)
192 on PC workstations running under the Windows 10 operating system. A custom LabView
193 (National Instruments) program was used to collect the behavioral data, also on a PC running the
194 Windows 10 operating system. Significant differences between groups were assessed with the
195 Mann-Whitney U test followed by p-value adjustment with False Discovery Rate when multiple
196 comparisons were made.

197 Predictability of pellet distributions was quantified using an across trial minimum distance
198 metric, which, for each pellet in a given distribution reports the minimum distance from that
199 pellet to all pellets in the immediately previous distribution. Relative entropy (RE) is equivalent
200 to Kullback-Leibler Divergence and was calculated as: $RE(P||Q) = \int P(j) \log(P(j)/Q(j))$ for
201 all points j in the current trial's probability density function (P) and the probability density
202 function calculated from all previous trials (Q). Prior to calculating the RE all distributions were
203 convolved with a smoothing function, which was an averaging filter of width = 1 cm. RE is
204 reported in bits.

205 For establishing optimal pellet acquisition sequences for each distribution we used a
206 genetic algorithm developed by Joseph Kirk: Fixed Start Open Traveling Salesman Problem -
207 Genetic Algorithm ([https://www.mathworks.com/matlabcentral/fileexchange/21198-fixed-start-](https://www.mathworks.com/matlabcentral/fileexchange/21198-fixed-start-open-traveling-salesman-problem-genetic-algorithm)
208 [open-traveling-salesman-problem-genetic-algorithm](https://www.mathworks.com/matlabcentral/fileexchange/21198-fixed-start-open-traveling-salesman-problem-genetic-algorithm)). Briefly, this algorithm starts from a

209 population of randomly generated paths that start at the entrance to the arena and travel to each
 210 pellet once. It then uses an iterative process wherein in each “generation” of solutions the fitness
 211 of every path in the population is evaluated; the objective function for fitness in this case is
 212 minimization of path length. The more fit (shorter) paths are selected, and each path’s sequence
 213 of pellet locations is modified (recombined with other paths or randomly changed, or “mutated”)
 214 to form a new generation. The new generation of candidate paths is then used in the next
 215 iteration of the algorithm. The algorithm can be terminated when either a maximum number of
 216 generations has happened or the path length reaches a small enough value.

217 Efficiency of foraging paths (Fig. 3a) was calculated as $fe = lo/la$, where lo is the
 218 optimal path length, la is the animal's path length, and fe is foraging efficiency.

219
 220 **Bayesian search.** For analyses conducted in Figure 5, we modeled rat behavior as a Bayesian
 221 search. Briefly, the search arena is divided into 2.8 cm squares resulting in a 40 x 80 grid of
 222 possible locations. This grid is then populated with the same pellet distributions that were used in
 223 the behavioral experiments. We start our analysis on day 10 of training, which provides an agent
 224 with up to the first 10 days of training data as a map of prior expectations regarding pellet
 225 locations (Fig. 5a). The expression for prior expectation of pellet location is given by:

$$226 \quad pe(x, y) = \Sigma_L^{-1}(x, y)/(t - L)$$

227
 228 Where t is the trial number, rw is the probability of a pellet being found at a given point, (x,y) ,
 229 over previous trials and pe is the resulting prior expectation from the previous pellet locations. L
 230 is based on the length of memory being used and is defined as $L = (t - md, 1)$, with md being
 231 memory depth in trials, with $md \geq 1$. To enforce the nearest-neighbor search strategy used by

232 rats, this map of prior expectations is discounted by linear distance from the agent, resulting in
 233 decreased likelihood to search first in areas that are located at large distances from the agent.
 234 This results in the following expression at a point, (x, y) within the grid of possible pellet
 235 locations:

$$236 \quad m(x, y) = pe(x, y) * ((\max(d) - d(x, y))/\max(d))$$

237
 238 Where d is the distance from the agent and m is the memory-based map of prior expectations for
 239 pellet location adjusted by distance from the agent. The agent also uses sensory information that
 240 decays with distance to update their expectation of the possible pellet location,

$$241 \quad s(x, y) = cr(x, y) * ((\max(d) - d(x, y))/\max(d))^{se}$$

242
 243 where s is the sensory density function and cr is a map with the current location of all pellets set
 244 to 1 and all other locations set to 0. The term se is an exponent that determines the rate of decay
 245 of sensory information with distance. These two sources of information are weighted and then
 246 summed to result in a map that guides the agent's next step in the search path.

$$247 \quad p(x, y) = s(x, y) * sw + m(x, y) * (1 - sw)$$

248
 249 Where p is the probability map, s is the sensory density function and m is the memory-based map
 250 of prior expectations for pellet location. The term sw is the weight given to sensory information,
 251 $\{sw \mid 0 \leq sw \leq 1\}$. The agent makes its next step along the vector to the maximum point of p .
 252 The agent is considered to have perfect target detection at their location, such that after the agent
 253 moves to a new location, if a pellet is at that location it is always detected and if no pellet is at
 254 that location the probability of a target at that site is updated to 0. To fit parameters for the
 255 Bayesian search, we used a 3-dimensional coarse grid of values for sw , se , and md . We found the

256 best fit for each animal in this grid and report these results in Figure 5.

257 For reported measures in Fig. 5f, $sa = (1 - (se/(SE)) + sw)/2$, where sa is sensory
258 acuity and SE is the set of values of se across all best fits for 12 animals, while

259 $mi = (max(pr\{md > 0\}) - mean(pr\{md \leq 3\}))/ (max(pr\{md > 0\}) - min(pr\{md > 0\}))$,
260 where mi is long-term memory usage and pr is the correlation of the agent's performance with
261 the animal's performance using md set to the indicated range of values.

262 **Software accessibility.** All software developed for analysis and generation of figures is available
263 at the Gire lab website and at Github.

264

265 **Results**

266

267 **Route planning revealed through controlling predictability of reward locations**

268 We adapted the probabilistic traveling salesman problem for experimental investigation
269 through the use of an automated system for precise, computer-controlled food pellet placement
270 within a large foraging arena (Fig. 1a). We divided a cohort of 12 rats into 3 equal groups that
271 foraged within environments of high, medium, and low food location predictability (Fig. 1b).
272 Animals in each group were tested across precisely replicated pellet placements (Fig. 1c) and all
273 placements used had equivalent optimal path lengths (Fig. 1d), as calculated through a genetic
274 algorithm solution to the traveling salesman problem for each pellet placement (see methods).
275 We generated sequences of pellet locations over days to create distributions that were extremely
276 well-predicted by prior experience as well as distributions that were unable to be anticipated
277 based upon prior pellet locations. To generate pellet placements with controlled levels of

278 predictability we quantified the between trial minimum distance for each pellet of a given
 279 distribution and all pellets of the previous trial's distribution and set this value to be low for the
 280 computer-generated set of locations used for predictable conditions and to be high for the
 281 unpredictable condition (Fig. 1e). The lower values for pellets in predictable distributions
 282 indicate that these pellets are in areas that are extremely close to where pellets were located on
 283 the previous trial, allowing animals to create an expectation over repeated searches. This is also
 284 demonstrated through a reduction of the relative entropy (a measure of surprise) of newly-
 285 encountered pellet distributions following multiple days of training for animals in high and
 286 medium predictability conditions. Animals could not develop such an expectation under low
 287 levels of predictability and relative entropy does not decrease with training for the unpredictable
 288 distribution (Fig. 1f). In all conditions, animals searched for an average of 7 pellets, with the
 289 precise number on a given trial unknown to the animal (Fig. 1g). This results in typically $7!$, or
 290 5,040 possible sequences of pellet acquisition, with most sequences being extremely sub-
 291 optimal. Examples of trajectories taken by animals on the first and last days of training
 292 demonstrate changes in search trajectories with learning (Fig. 1h). After training, all animals
 293 favored a small subset of near-optimal acquisition sequences (Fig. 2a), consistent with findings
 294 in non-probabilistic optimization across a number of species (Blaser & Ginchansky, 2012). We
 295 found that a simple nearest neighbor heuristic (in which rats solve the task by traveling to the
 296 next nearest pellet) achieved strong performance on this task, often comparable to that of
 297 optimized routes (Fig. 2b). Indeed, we found that animals achieved optimal performance only
 298 when the optimal solution was the same as a nearest neighbor approximating solution (Fig. 2c),
 299 suggesting that the rats employed the nearest neighbor heuristic to solve the task. Rats foraging
 300 in predictable environments were capable of employing a nearest neighbor strategy earlier during

301 training, though all animals, even those in unpredictable environments, did increase the use of
 302 nearest neighbor routes while foraging (Fig. 2d). However, animals in the highest predictability
 303 group were significantly more effective at ordering their search based on nearest neighbor
 304 relations of reward locations (Fig. 2e, error relative to a nearest neighbor search: 16.9 +/- 0.5 cm
 305 for most predictable, 22.1 +/- 2 cm for moderately predictable, and 20.8 +/- 1.4 cm for least
 306 predictable, n = 4 animals per predictability group, see methods for statistical tests used for all
 307 comparisons). Examples of optimal, nearest neighbor, and animal sequences of pellet acquisition
 308 for animals in highly predictable and unpredictable environments are shown in Fig. 2f.

310 **Predictable environments enable enhancement of search routes**

311 In our task, which involves probabilistic presence of pellets, this nearest neighbor
 312 search can be implemented through two different strategies: in a sensory-guided strategy
 313 animals use cues (odor or vision) to navigate towards the nearest detected target; in a memory-
 314 guided strategy animals use prior information to navigate towards the nearest, most likely
 315 locations of pellets. We next investigated which of the two alternative strategies might guide a
 316 nearest neighbor search within each level of uncertainty. Over training, animals across all
 317 predictability levels significantly increased their probability to travel to the nearest pellet
 318 during search (Fig. 2d). However, the number of days of training taken for this to occur was
 319 dependent upon the predictability of the pellet distribution (Fig. 2d; significant improvement
 320 on days 2-10 for highly and moderately predictable conditions, significant improvement not
 321 until days 10-15 for unpredictable conditions; $p < 0.05$ compared to day 1, $n=4$ for all groups).
 322 We found that animals searching in highly predictable environments were effective at
 323 enhancing the efficiency of their search across long distances (>40 cm) and learned to do this

324 relatively early in training (days 5-10). Those in moderately predictable environments also
 325 learned to increase the efficiency of their search tours but required more training to do so (days
 326 10-15), while those searching in unpredictable environments did not significantly increase the
 327 efficiency of their tours (Fig. 3a,b). As the unpredictable nature, or “surprise value” of the
 328 environment increased, the ability of animals to increase the efficiency of their search tours
 329 decreased (Fig. 3c, $R = -0.72$, $p < 0.008$, $n = 12$). These results suggest that based upon the
 330 predictability of the environment rats employ two different strategies to find the next nearest
 331 pellet – one in which tours can be efficiently narrowed towards straight line paths and another
 332 in which paths between rewards are necessarily circuitous (see Fig. 1h, lower panel for
 333 example tours after training).

334 In addition to supporting better-ordered search routes (Fig. 2d-f) and efficient paths to
 335 the nearest target from farther away (Fig. 3a-c), predictable distributions also enabled rats to
 336 enhance the speed of their travel between rewards. During training, the speed of the trajectories
 337 taken between pellets increased the most quickly for animals operating in the most predictable
 338 environments, though all animals eventually learned to decrease time between rewards by
 339 increasing speed (Fig. 3d). Time spent pausing (speed < 1 cm per second) and number of
 340 pauses per second did not significantly change with training (Fig. 3e,f), suggesting consistent
 341 motivation to perform the task across all animals.

342 **Analyzing shifting weightings between sensory- and memory-dominated strategies**

344 We next sought to more precisely quantify the role of sensory information and memory
 345 in the navigation strategies used by animals under varying levels of uncertainty. To perform this
 346 analysis we simulated animal behavior by developing an agent that searched through foraging

space using multiple free parameters related to exploratory and exploitative search characteristics (Elazary & Itti, 2010; McNamara et al., 2006). These parameters include the length of memory for the prior, the distance over which sensory signals from the pellets are detected, and the relative weighting of sensory and memory terms. We allowed these parameters to vary on a multidimensional grid and analyzed goodness of fit to actual animal performance as the correlation between trial-by-trial performance of the simulated searcher and the animal (Fig. 5 and see methods). As expected, searches with long-range, noiseless sensory information lead to a perfect nearest neighbor search and do not correlate well with animal behavior (Fig. 5b) since rats do not have access to perfect information and need to use local sensory information or learned locations to navigate (see Fig. 4 for an examination of possible sensory cues used for this task). Similarly, searches with only a memory term also do not correlate well with actual behavior (Fig. 5b). Consistent with animals under different levels of uncertainty using diverse search strategies, we found that any set of a wide range of parameters applied uniformly to all animals resulted in only moderate correlation with actual behavior (Fig. 5c). We next allowed parameters to vary individually for each animal. While this approach will trivially result in a better fit due to the increased number of free parameters (Fig. 5b-c, $p < 0.01$; $n = 12$), we used the values of parameters obtained for these individual fits to examine the contribution of sensory and memory input to the simulated search that best matched each animal's performance. When varying the length of memory used by the searcher we found that simulated searches across the most predictable distributions benefited from increased memory with an increase in correlation to actual animal performance when the simulated searcher had access to cumulative memory of previous searches (predictable, single trial memory: $R = 0.12 \pm 0.05$; cumulative memory $R = 0.66 \pm 0.03$; $p < 0.05$; $n = 4$). Searches across moderately predictable and unpredictable

distributions did not show a significant increase in correlation with animal behavior with increased memory (Fig. 5d). Consistent with these results, the impact of shuffling prior distributions on agent performance was directly related to the predictability of the data set (Fig. 5e). To quantify the impact of sensory input on these searches we combined the weighting given to sensory input with the distance from which each agent could detect a target to create a measure of sensory acuity for each simulated agent (see methods). This measure was well correlated with increasing relative entropy of the training set, suggesting that animals increased sensory acuity under uncertainty (Fig. 5f, left panel; $R = 0.8469$; $p = 0.005$). We also used the length of memory for the best match to animal behavior to create a metric for long-term memory usage (see methods). We found a significant inverse correlation between relative entropy and long-term memory usage (Fig. 5f, right panel; $R = -0.7252$; $p = 0.0076$), suggesting that as the training set became more predictable animals relied more on long-term memory. Our results are consistent with a Bayesian search where searchers adaptively shift the weightings given to various locations (and thus, their likelihood to travel to these locations) based on their relative weightings of sensory and memory terms. For example, a searcher may shift the weighting of a given location based on being rewarded there many times in the past (exploitative, memory-guided strategy) or it may shift the weighting based on sensing cues emanating from a given location (exploratory, sensory-guided strategy).

388

389 **Discussion (1,402 words)**

390

391 Animals make use of appropriate cognitive strategies and behaviors to solve the many
392 problems they are faced with during self-guided behaviors such as foraging (Marewski & Link,
393 2014). It is known that when animals are introduced to new environments with multiple food
394 locations they may continually explore and sample the different options, or they may exploit a

single, most profitable option (Krebs, 1978). However, it is not fully understood how animals balance exploratory behaviors against exploitative behaviors (Gupta et al., 2006). Our study revealed that rodents make use of their prior knowledge of the predictability of an environment to determine the extent that they rely on sensory cues during their foraging bouts. Our results are consistent with a strategy that increases sensory acuity and reduces memory load in direct relation to the level of uncertainty in an environment (Fig. 6). This increased reliance on sensory input allows animals searching across unpredictable environments to employ an effective nearest neighbor strategy with nearly the same efficacy as animals that are operating in highly predictable environments, although due to the short-range nature of sensory cues a sensory-guided strategy fails at long distances and animals are unable to increase the efficiency of foraging trajectories over these distances (Fig. 3). Conversely, animals operating in predictable environments reduce their reliance on sensory input in favor of stereotyped and efficient searches based on long-term memory, which allows them to enhance search tours over long distances. In short, in a sensory-dominated strategy animals approach the nearest sensed pellet, while in a memory-dominated strategy animals approach the nearest remembered location, enabling more efficient, planned routes to emerge. This result is consistent with the finding that humans integrate information from different sensory modalities and dynamically give greater weight to the modality that provides the stronger, most well-defined estimate (Ernst & Banks, 2002). Taken together, these results suggest that animals assess the predictability of an environment to select appropriate strategies to allocate cognitive resources between sensory processing and memory while solving complex natural problems.

While it is difficult for animals to rapidly learn efficient paths for collecting rewards in the unpredictable environment, optimal paths in this environment are not more complex than

418 those in predictable environments, as shown in figure 1D. Indeed, animals in unpredictable
419 environments do optimize their foraging behavior after many sessions, achieving a roughly equal
420 ability to perform a nearest-neighbor solution to the task (see Fig. 2d). They may learn a general
421 understanding of where pellets have never been found (such as along the boundaries of the
422 arena) and may focus their search to the center of the arena in order to maximize getting close
423 enough to pellets to then use sensory guidance to approach the reward locations (examples in
424 Fig. 1b,h). This suggests that while animals have a diminished, imperfect ability to rapidly learn
425 efficient paths in unpredictable environments they are still capable of improving their foraging
426 strategy, perhaps through a combination of coarse predictions and enhanced sensory guidance.

427 The differential weighting of sensory cues, specifically odor cues, is expected when the
428 turbulent nature of odor plumes in natural environments is taken into account. Odor-guided
429 searches are notoriously difficult due to the sparse and intermittent nature of odor plumes
430 (Vickers, 2000). The ability of rodents to form internal representations of their environment
431 could allow them to apply learned spatial information to dynamic environments, creating a map
432 that would act to lessen the cognitive load required to use the complex sensory cues in odor
433 plumes and greatly increase the effectiveness of odor-guided searches. So it follows that rodents
434 would prefer to use a strategy that relies less on olfactory cues when instead they could navigate
435 using the cognitive map of their familiar environment. This is in line with our results suggesting
436 that under unpredictable conditions rats do not efficiently navigate to the next closest pellet when
437 it is more than 40 cm from their current location (Fig. 3a). Previous research suggests that 40cm
438 is close to the threshold of rodents' ability to gain a directional benefit from the sparse odor cues
439 emanating from an odor source (Gire et al., 2016; Liu et al., 2020). This difficulty is increased
440 when rats have been trained on unpredictable environments and are unable to construct strong

441 expectations of pellet location. Since there is no underlying structure of where pellets can be
442 found that animals in the unpredictable environment can learn over time, the low weighting
443 given to the memory terms in the Bayesian model reflects animals' discounting of information
444 that will not be as useful as increasing their reliance on sensory cues. Animals then take
445 advantage of the sensory cues emanating from food locations by increasing their weighting,
446 which is in line with the results from our Bayesian model (Fig. 5). Monitoring the trajectories of
447 the rats allowed us to also determine that rats traveled in much more efficient paths when they
448 were navigating under conditions of high predictability. This suggests that they are able to
449 navigate directly to where pellets are located without having to resort to behaviors indicative of
450 searching for olfactory cues, which typically result in more circuitous search trajectories (see
451 Fig. 1h).

452 Optimizing travel paths during navigation is a notoriously difficult problem to solve,
453 especially when one considers the complexity of the traveling salesman problem. One must
454 determine the shortest path between multiple locations in order to travel efficiently and conserve
455 the most energy or increase the rate of reward per unit time. This problem is extremely difficult
456 to solve optimally as the complexity of the problem scales unfavorably with the number of
457 targets that must be visited. In our task, this problem is even more complex due to the fact that
458 animals only have probabilistic information about whether food pellets will be present at target
459 locations. While not optimal, simplifying heuristics enable solutions to such complex
460 optimization problems to be reached in relatively short periods of time. Nearest neighbor tours
461 are a common strategy used to solve the traveling salesman problem (Johnson, 1990; Tsai et al.,
462 2004). Under this strategy, the agent simply travels to the next nearest target location until all
463 targets have been visited. While not optimal, this approach is computationally simple, resulting

464 in rapid solutions with time to solve scaling well with task complexity. Our results suggest that
465 animals adopt a nearest neighbor strategy to procure all of the pellets; however, the degree to
466 which the strategy resembles a perfect nearest neighbor strategy depends on the predictability of
467 the environment. Animals trained in a predictable environment select a strategy that highly
468 resembles a nearest neighbor search earlier on in training (Fig. 2), which allows them to more
469 effectively exploit pellet locations and increase efficiency (Fig. 3a) and speed (Fig. 3d) of their
470 routes. In contrast, animals trained in unpredictable environments select a strategy that resembles
471 a nearest neighbor search much later in training (Fig. 2d). These differential time courses could
472 reflect the time necessary to train the underlying memory or sensory networks in the brain, with
473 sensory training requiring a longer training period.

474 The novel, fully-automated foraging arena we designed allows for new ways to study the
475 balance between exploration and exploitation. Using an automated, moving pellet dispenser
476 allows for food rewards to be placed in an unlimited number of different locations throughout the
477 foraging arena. This allows us to instantaneously change any location in the arena into a reward
478 location. Instead of being confined to defined locations, such as fixed near a feeder, we are able
479 to create many different distributions of where food can be found, mimicking distributions that
480 might occur in a more naturalistic setting. By combining this automated arena with computer-
481 generated reward distributions we can also scale the difficulty of the task to address specific
482 research questions. This allows us to study more complex behaviors that current experimental
483 paradigms are not equipped to adequately explore. Through computer-aided creation of reward
484 location sequences our new approach also supports direct testing of algorithms that could be used
485 to perform self-guided optimization. This task also integrates extremely well with new advances
486 in automated behavioral tracking (Nath et al., 2019). Finally, the self-guided nature of our task

487 allows for future studies to elucidate neural mechanisms underlying complex behaviors, such as
488 route optimization. Since animals trained on this task are not explicitly shaped or instructed on
489 how to best perform, we are able to study how the brain changes as animals develop solutions to
490 complex, natural problems.

491

References

- Agarwal, G., Stevenson, I.H., Berényi, A., Mizuseki, K., Buzsáki, G., & Sommer, F.T. (2014). Spatially distributed local fields in the hippocampus encode rat position. *Science*, *344*(6184), 626-630.
- Auh, S. & Menguc, B. (2005). Balancing exploration and exploitation: The moderating role of competitive intensity. *Journal of Business Research*, *58*(12), 1652-1661.
- Blaser, R. E. & Ginchansky, R. R. Route selection by rats and humans in a navigational traveling salesman problem (2012). *Animal Cognition*, *15*(105), 239–250
- Dolan, R.J. & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*, 312-325.
- Drea, C. M., & Carter, A. N. (2009). Cooperative problem solving in a social carnivore. *Animal Behaviour*, *78*(4), 967–977.
- Elazary, L., & Itti, L. (2010). A Bayesian model for efficient visual search and recognition. *Vision Research*, *50*(14), 1338–1352.
- Ernst, M.O. & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429-433.
- Fuentes, G. E. A., Gress, E. S. H., Mora, J. C. S. T., & Marín, J. M. (2018). Solution to travelling salesman problem by clusters and a modified multi-restart iterated local search metaheuristic. *PLOS ONE*, *13*(8), e0201868
- Gire, D.H., Kapoor, V., Arrighi-Allisan, A., Seminara, A., & Murthy, V.N. (2016). Mice develop efficient strategies for foraging and navigation using complex natural stimuli. *Current Biology*, *26*(10), 1261-1273.
- Gupta, A.K., Smith, K.G., & Shalley, C.E. (2006). The interplay between exploration and exploitation. *Academy of Management Journal*, *49*(4), 693-706.
- Johnson, D. S. (1990). Local optimization and the traveling salesman problem. In M. S. Paterson (Ed.), *Automata, Languages and Programming* (Vol. 443, pp. 446–461). Berlin/Heidelberg: Springer-Verlag.
- Kenward, B., Weir, A. A. S., Rutz, C., & Kacelnik, A. (2005). Tool manufacture by naive juvenile crows. *Nature*, *433*, 121.
- Kralik, J. D., Xu, E. R., Knight, E. J., Khan, S. A., & Levine, W. J. (2012). When less is more: Evolutionary origins of the affect heuristic. *PLOS ONE*, *7*(10), e46240
- Kramer, D.L. & Weary, D.M. (1991). Exploration versus exploitation: A field study of time allocation to environmental tracking by foraging chipmunks. *Animal Behaviour*, *41*(3), 443-449
- Krebs, J.R., Kacelnik, A., & Taylor, P. (1978). Test of optimal sampling by foraging great tits. *Nature*, *275*, (27-31).

- 542
543 Leipälä, T. (1978). On the solutions of stochastic traveling salesman problems. *European Journal of*
544 *Operational Research*, 2(4), 291–297
545
- 546 Marewski, J.N. & Link, D. (2014). Strategy selection: An introduction to the modeling challenge.
547 *WIREs Cognitive Science*, 5, 39–59.
548
- 549 Marsh, B. (2002). Do animals use heuristics? *Journal of Bioeconomics*, 4(1), 49–56.
550
- 551 McNamara, J. M., Green, R. F., & Olsson, O. (2006). Bayes' theorem and its applications in animal
552 behaviour. *Oikos*, 112(2), 243–251.
553
- 554 Mehlhorn, K., Newell, B.R., Todd, P.M., Lee, M.D., Morgan, K., Braithwaite, V.A., Hausmann,
555 D., Fiedler, K., & Gonzalez, C. (2015). Unpacking the exploration-exploitation tradeoff: A
556 synthesis of human and animal literatures. *Decision*, 2(3), 191–215.
557
- 558 Mukherjee, A., Panigrahi, G., Kar, S., & Maiti, M. (2019). Constrained covering solid travelling
559 salesman problems in uncertain environment. *Journal of Ambient Intelligence and Humanized*
560 *Computing*, 10(1), 125–141.
561
- 562 Nath, T., Mathis, A., Chen, A.C., Patel, A., Bethge, M., & Mathis, M.W. (2019). Using DeepLabCut
563 for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, 14, 2152–
564 2176.
565
- 566 Pedigo, S.F., Song, E.Y., Jung, M.W., & Kim, J.J. (2006). A computer vision-based automated
567 Figure-8 maze for working memory test in rodents. *Journal of Neuroscience Methods*,
568 156(1–2), 10–16
569
- 570 Percus, A. G. & Martin, O. C. (1999). The stochastic traveling salesman problem: Finite size
571 scaling and the cavity prediction. *Journal of Statistical Physics*, 94(5/6), 739–758.
572
- 573 Prusky, G. T., Harker, K. T., Douglas, R. M. & Whishaw, I. Q. (2002). Variation in visual acuity
574 within pigmented, and between pigmented and albino rat strains. *Behav. Brain Res.* **136**,
575 339– 348.
576
- 577 Prusky, G. T., West, P. W. & Douglas, R. M. (2000). Behavioral assessment of visual acuity in
578 mice and rats. *Vision Res.* **40**, 2201–2209.
579
- 580 Raman, V., & Gill, N. S. (2017). Review of different heuristic algorithms for solving travelling
581 salesman problem. *International Journal of Advanced Research in Computer Science*, 8(5),
582 423– 425.
583
- 584 Slotnick, B. (2001). Animal cognition and the rat olfactory system. *Trends in Cognitive Sciences*,
585 5(5), 216–222.
586
- 587 Tsai, C.-F., Tsai, C.-W., & Tseng, C.-C. (2004). A new hybrid heuristic approach for solving large
588 traveling salesman problem. *Information Sciences*, 166(1), 67–81.
589
- 590 Vickers, N.J. (2000). Mechanisms of animal navigation in odor plumes. *The Biological Bulletin*,
591 198(2), 203–212.

- 592
593 Wall, S. B. V., & Balda, R. P. (1977). Coadaptations of the Clark's Nutcracker and the Pinon Pine
594 for efficient seed harvest and dispersal. *Ecological Monographs*, 47(1), 89–111.
595
596 Zhang, S. & Manahan-Vaughan, D. (2015). Spatial olfactory learning contributes to place field
597 formation in the hippocampus. *Cerebral Cortex*, 25(2), 423–432.
598
599 Zhang, Q., Yang, X., Li, P., Huang, G., Feng, S., Shen, C., & Lu, T. J. (2015). Bioinspired engineering of
600 honeycomb structure – Using nature to inspire human innovation. *Progress in Materials Science*,
601 74, 332–400
602
603
604

Figure Legends

Fig. 1: A computer-controlled probabilistic traveling salesman task enables direct tests of behavioral strategies under uncertainty.

a) (top) A large, automated arena with a rat shown for scale. **(bottom)** The temporal structure of a typical trial. **b)** Rats forage for pellets in highly predictable (**left**), moderately predictable (**center**), and actively randomized (**right**) pellet placements. Placements are shown across all trials (20 days, 3 trials per day). **c)** The automated system allows for reproducible pellet placement across animals. From the top to bottom of the matrix correlation coefficients are shown for two different predictable distributions and the single unpredictable distribution. **d)** Pellet distributions from each placement shown in panels b) and c) have equivalent optimal path lengths. **e)** Example histograms are shown for the most predictable (black) and least predictable (gray) distributions that were tested. Vertical colored lines show the mean for the predictable (blue) and unpredictable (red) distributions. The distributions for all animals are plotted as colored circles, with color corresponding to across trial minimum distance. **f)** Relative entropy for each predictability grouping (high - blue; medium - purple; and randomized - red) across sessions of training. Higher values indicate higher entropy. **g)** Average number of pellets per trial for each predictability level. **h)** Examples of routes taken by rats on the first trial of the first day (top panels) and after 20 days of training (bottom panels). Color shifts from cyan to yellow as each animal's trajectory progresses.

Fig. 2: Search performance approaches a nearest neighbor heuristic after experience with reward locations

a) Average distance per pellet. Rats acquire pellets in a sequence that is extremely efficient (red lines) compared to a random sampling of all possible sequences (blue bars). Predictability decreases from top to bottom. **b)** Performance of a nearest neighbor strategy on all distributions tested in this study when compared to the optimal path length. Dashed lines represent 10, 20, and 30% above optimal. **c)** Animal performance on trials in which a nearest neighbor search is optimal vs. trials in which a nearest neighbor search is sub-optimal. **d)** The probability that rats in each predictability group acquire the nearest pellet during search increases during training for all groups. **e)** Scatter plot showing the relation between predictability of distribution (x axis) and difference between animal acquisition sequence and nearest neighbor sequence (y axis) for all 12 animals. **f)** Example of optimal and nearest neighbor pellet acquisition sequences, and the actual sequences and trajectories taken by animals. For the right panels, color shifts from cyan to yellow as the animal's trajectory progresses and from dark to light blue as the pellet acquisition sequence progresses.

Fig. 3: Predictability supports increased route efficiency.

a) Animals searching in predictable environments increase efficiency with training (see methods for efficiency metric). Efficiency was measured on paths to rewards that were located more than 40 cm away and were assessed on day 1 and then on blocks of 5 days until day 20. **b)** Animals in both predictable groups significantly increased the efficiency of their search routes on the last block of training when compared to the first day. **c)** Efficiency of search routes measured on the

last block of training (days 15-20) show a strong negative correlation to the unpredictability of the foraging environment, here measured as the cross trial minimum pellet distance (see methods). **d)** All animals increase speed during training. Average speed was taken without including pauses. **e)** Animals spend a small amount of time pausing during the task and this does not significantly change with training. **f)** The number of pauses per route as a function of training.

Fig. 4: Sensory cues are local.

a) Top: Experimentally determined spread of odor in the foraging arena (see methods). **Bottom:** Calculated size of a pellet necessary for it to be visible for a foraging rat under bright, broad-spectrum lighting conditions with high contrast, based on reported values for rat visual acuity. The dashed red line indicates the actual size of the pellets used (and thus the distance for detection under ideal conditions). All experiments in the current study were done under dim red light using pellets matched in color to the arena floor, further limiting the range for visual detection. **b)** Estimated best-case pellet detection distances for olfactory (cyan) and visual (red) sensory cues. Due to both the dim, red lighting conditions and the lightly odorized pellets actual detection distances are likely to be much smaller. **c)** The entire time course of odor for one mapping experiment (approximately 180 minutes) used to establish the distribution in panel a. As the sensor is moved closer to the source (later in the experiment) odor fluctuations become much larger. **d)** A grid of mean odor intensity values that were sampled during the experiment and convolved with a gaussian function to create the estimated odor density function in panel a. Odor sensor activation over time from the indicated locations (1,2 and 3) is shown to the left of the grid.

Fig. 5: Modeling behavior as a Bayesian search with adaptive sensory acuity and memory depth explains performance under uncertainty.

a) Examples of prior distributions accumulated over all trials for one predictable and one unpredictable set of pellet locations. **b)** Correlation to animal performance of models with parameters emphasizing sensory (S) or memory (M) guidance or an adaptive model (A) individually fit to each animal. **c)** Correlation of agent's search performance with animal behavior when using parameters fit to other animals (All others) or the best fit to that specific animal (Best fit). The best fit is significantly better than the fits from other animals ($p = 0.0043$, $n = 12$). **d)** Correlation between animal behavior and a Bayesian search with either single trial memory (1) or best performance with cumulative memory (C). **e)** Performance ratio (Path length with priors from different distributions / Path length with correct prior) for all animals plotted as a function of the across trial minimum distance for the distributions presented to each animal (significant correlation: $R = -0.85$, $p = 0.0004$). A higher value for the performance ratio indicates longer path length with a shuffled prior. Agents searching with unpredictable distributions (red) show identical performance regardless of the prior used. **f) (left)** Sensory acuity based on the best fit search parameters vs. relative entropy based on the distributions that animals have experienced. **(right)** Long-term memory usage vs. relative entropy of pellet distributions encountered.

694 **Fig. 6: Schematic of two strategies selected to solve the probabilistic traveling salesman**
695 **task.** A schematic of the main results, showing that animals adaptively change the strategies used
696 for a search depending upon the level of uncertainty of the environment, here depicted as a
697 spectrum from red (uncertain) to blue (predictable).
698
699











