The traveling salesrat: insights into the dynamics of efficient spatial navigation in the rodent

The traveling salesrat: insights into the dynamics of efficient spatial navigation in the rodent

Laurel Watkins de Jong[1,2]**, Brian Gereke**[2]**, Gerard M Martin**[3]
and Jean-Marc Fellous[2,4,5]

[1] Graduate Interdisciplinary Program in Neuroscience, University of Arizona, Tucson, AZ 85724, USA
[2] Computational and Experimental Neuroscience Laboratory, University of Arizona, Tucson, AZ 85724, USA
[3] Department of Psychology, Memorial University of Newfoundland, St John's, Newfoundland and Labrador, Canada
[4] Department of Psychology and Applied Mathematics, University of Arizona, Tucson, AZ 85724, USA

E-mail: fellous@email.arizona.edu

Abstract

Rodent spatial navigation requires the dynamic evaluation of multiple sources of information, including visual cues, self-motion signals and reward signals. The nature of the evaluation, its dynamics and the relative weighting of the multiple information streams are largely unknown and have generated many hypotheses in the field of robotics. We use the framework of the traveling salesperson problem (TSP) to study how this evaluation may be achieved. The TSP is a classical artificial intelligence NP-hard problem that requires an agent to visit a fixed set of locations once, minimizing the total distance traveled. We show that after a few trials, rats converge on a short route between rewarded food cups. We propose that this route emerges from a series of local decisions that are derived from weighing information embedded in the context of the task. We study the relative weighting of spatial and reward information and establish that, in the conditions of this experiment, when the contingencies are not in conflict, rats choose the spatial or reward optimal solution. There was a trend toward a preference for space when the contingencies were in conflict. We also show that the spatial decision about which cup to go to next is biased by the orientation of the animal. Reward contingencies are also shown to significantly and dynamically modulate the decision-making process. This paradigm will allow for further neurophysiological studies aimed at understanding the synergistic role of brain areas involved in planning, reward processing and spatial navigation. These insights will in turn suggest new neural-like architectures for the control of mobile autonomous robots.

Introduction

The best way to go from point A to point B is on a straight line. What is the best way to go through N points? This question is the basis of the classic traveling salesman/salesperson problem (TSP), in which an agent has to go through N different cities using the shortest path without ever revisiting any city. This

problem is NP-hard, in that there is no known algorithm that can efficiently solve the problem in polynomial time. There is not even a proof that such an algorithm exists. Decades of research have produced algorithms and heuristics that have achieved trade-offs between the variables involved in the problem (Gutin and Punnen 2002, Applegate 2006). Many of them have contributed to significant advances in theories of complexity and decision-making. Some of these approaches have been implemented in the context of autonomous robotics

systems, efficient spatial navigation (Cardema *et al* 2004, Blum *et al* 2007) and in industrial process scheduling (Bagchi *et al* 2006). Little is known, however, about the way the brain solves problems of this sort. With the recent advances made in understanding the neural substrates of decision-making and spatial navigation in humans, monkeys and rats, the time is ripe to revisit the issue from a neuroscience point of view.

Humans can find near-optimal solutions to computer-based versions of the TSP, but they typically do this intuitively, using perceptual information, and there is a large variability of strategy from individual to individual (Tenbrink and Wiener 2009). Gestalt factors such as aesthetics or symmetry of the city layout enter into play and correlate significantly with the mathematical optimality of the routes (MacGregor *et al* 2004, Vickers *et al* 2006). These findings indicate that spatial optimization may be a very basic feature of the nervous system, occurring as early as the sensory areas. It is still difficult to study the neural basis of the TSP moment-to-moment decision-making process in humans, as it would involve invasive recording procedures available only in rare cases. An animal model of the TSP in which the neural substrate of spatial representations can be accessed would be a significant advance in the field.

There have been many behavioral studies of spatial navigation in bees, ants and other insects (Marshall *et al* 2009). Most have shown that their collective behavior yields near-optimal and highly adaptable navigation. Studies of optimal spatial navigation in vertebrates have been scarcer, probably because of large individual differences in strategies and cost functions. Chimpanzees are able to find near-optimal TSP solutions with 18 baited locations (Menzel 1973). Interestingly, when the type of bait included preferred and non-preferred foods, animals changed their route to primarily include preferred food, clearly demonstrating that the reward value, not just location, was important to the animals. Similar conclusions were reached with baboons (Noser and Byrne 2010). Studies in vervet monkeys showed that path planning involved the consideration of at least three upcoming spatial targets, demonstrating that their strategy was not simply to go to the next closest location (Gallistel and Cramer 1996, Cramer and Gallistel 1997). Similarly, in a perceptual version of the task, macaques have been shown to optimize their eye movements to minimize distance between visual targets (Desrochers *et al* 2010). The ability to spatially 'think ahead' may be absent in some species such as pigeons which tend to use the nearest-neighbor strategy (Gibson *et al* 2007, Miyata and Fujita 2010). In a seminal study, rats were required to navigate through six spatial locations in a small arena (Bures *et al* 1992). After ten trials, the animals adopted a near-optimal route. While reminiscent of the TSP, this task differed in several significant ways: (1) reward was not given until the sixth city had been visited; (2) most city configurations were symmetrically positioned, and their maximum distance was small (on the order of 20 cm), not requiring much effort; and (3) animals were given the same city configuration for ten trials each day over six days, which clearly engaged long-term memory components that are not typically considered relevant to the decision-making processes involved in the TSP.

Spatial navigation in the rodent relies on a well-known set of brain structures, including the hippocampus and entorhinal cortex (Andersen *et al* 2007, Mizumori 2008). A long history of research has shown that these structures contain a complex set of neurons that are sensitive to head orientation, visual cues, spatial context, spatial location and task demands. These structures, together with cortical structures such as the parietal and the prefrontal cortices, compute the correct trajectory required to reach a target goal (Ainge *et al* 2007, Hok *et al* 2007). In addition, there is some evidence that the firing of these cells may be modulated by the location and/or availability of rewards (Dupret *et al* 2010). There is also evidence that the firing fields of these cells are modulated by dopaminergic projections from the ventral tegmental area (VTA), a brain region known to process rewards (Schultz 2010, Martig and Mizumori 2011). Little, if anything, is known about the manner in which these various brain areas cooperate to compute a near-optimal route.

We propose a new rodent task reminiscent of the TSP. We show that rats can naturally find short routes after a few trials within a single session. We show that their spatial navigation decisions depend on both spatial and reward cues. In the spatial domain, their decision depends on distance and orientation. In the reward domain, they are sensitive to the magnitude of rewards. We also show that they dynamically replan their route to ignore locations that have unexpectedly lost their rewarding value.

Methods

Animals

Data were obtained from nine Brown Norway–Fisher 344 hybrid rats approximately 8–14 months old. All rats were kept in separate cages in the same colony room operating under a reverse 12:12 h light cycle. Experiments were performed during the dark phase of this cycle. Upon arrival, rats were handled for 30 min per day for four days while being kept on free food to determine a baseline weight. After four days, the rats were exposed to the experimental room for 30 min per day while being food deprived. Once they reached 85% of their baseline weight, rats were pretrained for 30–45 min per day to exit a starting box and eat food pellets from small cups positioned on an open field arena. Pretraining lasted up to two weeks. Rats were always run on the orientation and distance versus reward experiments, where they were only allowed to visit one set of cups, before the fixed-N or variable-N experiments where they were allowed to visit multiple reward locations.

Behavioral apparatus

The open field arena was a round black table 152 cm in diameter with a 33 cm high wall surrounding the circumference (figure 1(A)). The feeder cups, created from plastic weighing boats, were 4×4 cm^2 wide and elevated 3 cm so that the rats could not see their contents unless they were very close. The start box was designed to control orientation. The box was black, 12×20 cm^2 wide and 28 cm tall and was elevated by
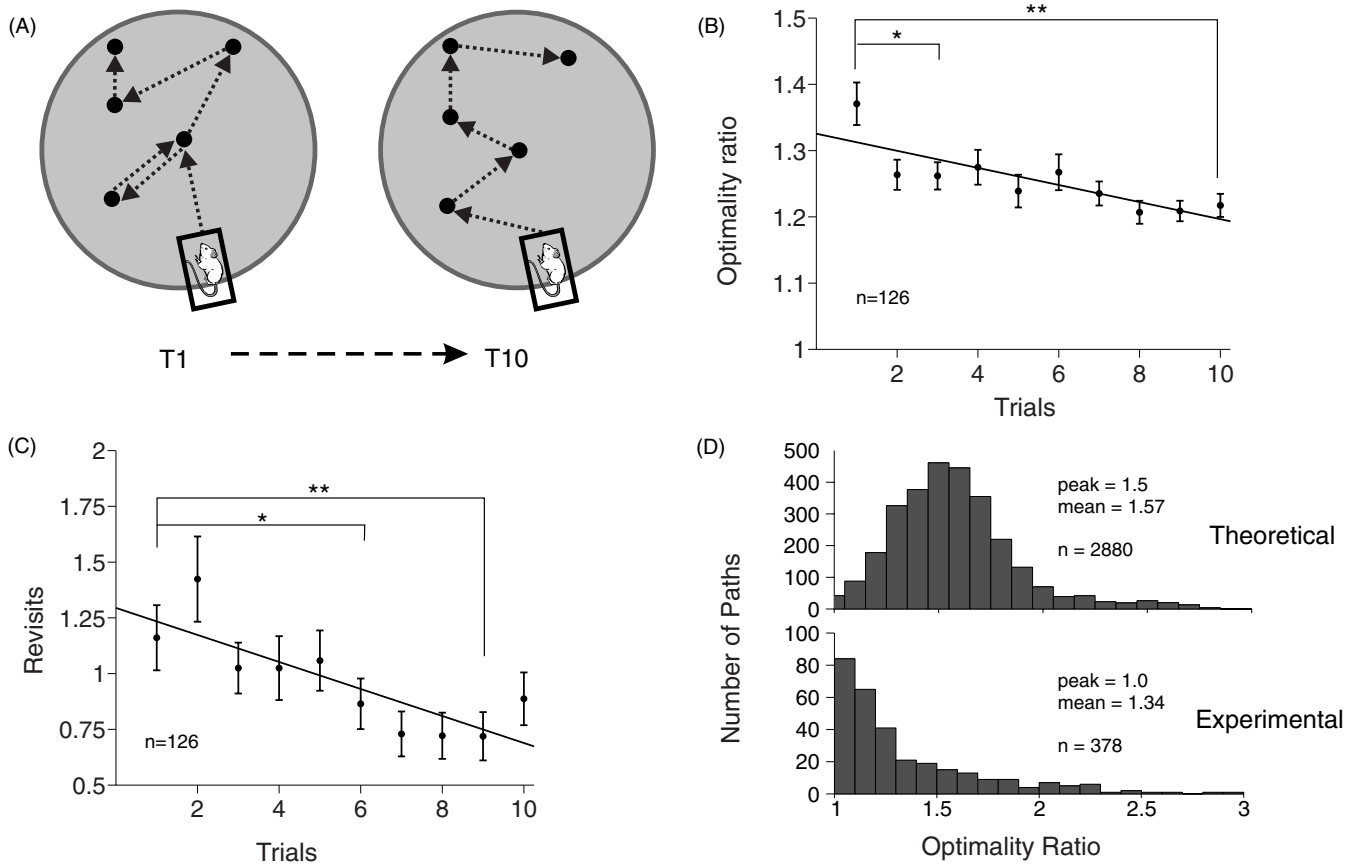
Figure 1. Rats find near short solutions to the 'fixed-N' TSP. (A) Cartoon illustration of the five-city TSP configuration for trials T1 and T10. (B) Average optimality ratio for paths with no revisits across rats and configurations. Optimality is taken as the ratio of the actual path to the shortest straight-line path possible to visit all the target locations. The '*n*' value corresponds to the number of configurations summed across rats. (C) Average number of revisits per trial for the five-city configurations. There is no significant difference ($p = 0.06$) between the number of revisits on the first and last trials. A total of 126 configurations were tested. (D) *Top*: theoretical distribution of optimality ratios of all possible straight-line routes for the 24 five-city configurations used in our experiments. *Bottom*: distribution of optimality ratios for routes chosen by rats for the last three trials of the same configurations as in (B) and (C). The *n* corresponds to the number of paths.

1 cm to protect the rats' tails. It included a vertical guillotine-style door. With the door removed, a rat could exit the box through an 8×6 cm^2 cutout in the front of the box. The cutout was small enough for a rat to exit by walking straight out. Rats wore a reflective strip of Velcro positioned just behind their forepaws that could be tracked by an overhead camera. The room contained shelving with laboratory supplies, and a door ~3 ft from the arena. While no attempt was made to strictly control for distal visual cues in the room, all major cues (e.g. door, shelving) were kept constant. The high walls at the periphery of the arena minimized the influence of local cues outside the arena. At the end of any given trial, a large felt cylinder was lowered over the rat. The cylinder was then used to return the rat to its original starting location, where the starting box was lowered back over the rat. This procedure was designed so that the experimenter never physically handled the rat during the experiment to minimize stress to the animal.

Fixed-N experiment. Rats were presented with spatial configurations containing five reward locations (figure 1). Cups were baited with 20 mg food pellets and could be located at any of 21 possible locations evenly distributed on the table

so that the minimal and maximal distances between any two cups were 25 and 120 cm, respectively. Configurations were designed so that only one cup could be located along one of the walls of the arena. The starting position of the box was always located against the arena wall, and was chosen to maximize the distance to the nearest cup. The starting box was always oriented at $0°$, toward the center of the arena. Rats were given ten trials to learn each configuration, and 90 s to complete each trial. This time constraint did not place additional demands on the task, since rats typically completed each trial in less than 30 s. If a rat timed out three times before completing ten trials, a new configuration was presented. The experiment terminated if a rat failed to complete three configurations and was resumed the next day. Rats were exposed to four configurations per day until they completed six full days. For this experiment, three rats were tested on 24 separate configurations and four rats were tested on a reduced three-day, 12-configuration protocol.

Variable-N experiment. The setup for the variable-N experiment was the same as for the fixed-N experiment except that configurations could consist of four–nine cities (figure 4). Four rats were tested on a three-day protocol.
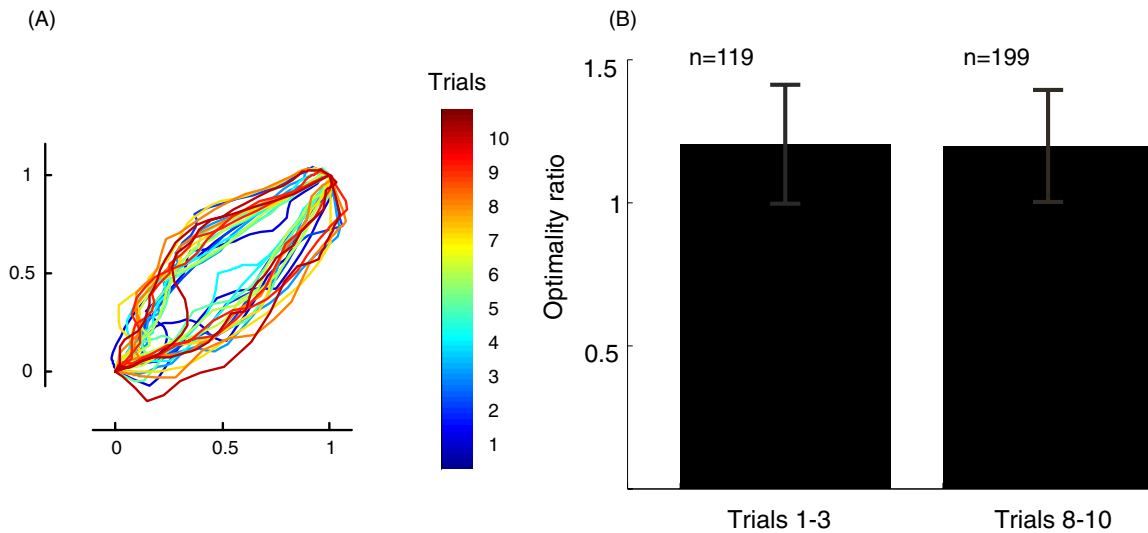
Figure 2. Trajectory minimization. (A) Sample trajectories taken by the rat. For each pair of consecutively visited cities within a configuration, blue colors represent paths taken during the initial trials and red colors are paths taken during the last trials. City locations and paths have been rescaled so that one city occupies (0,0) and the other location (1,1). This sample was taken for a five-city configuration. (B) Overall analysis across four–nine cities (all data in figure 1). The left bar is the average of the first three trials, the right bar is the average of the last three trials for each configuration and the *Y*-axis is the ratio of the path length to the minimum path possible (sqrt(2) after normalization). Only paths smaller than the Mahalanobis distance (here 2) have been included in the analyses, so as not to bias the analyses with trajectories that were taken when the animal was exploring the maze. The '*n*' corresponds to the number of trials.

(This figure is in colour only in the electronic version)

Two rats repeated the protocol for a total of six days. On day one, rats were run on three seven-city and three four-city configurations. Day two consisted of three eight-city and three five-city configurations. Day three consisted of three nine-city and three six-city configurations. A 15 min rest session was always inserted between changes in the number of cities. The rest session involved placing the rat in a covered container that was placed at the center of the arena. The configurations were identical across rats.

Orientation experiment. The configuration layout consisted of two reward locations (figure 5) located 60 cm from the starting location and ±30° from the axis which bisected the arena and passed through the starting location. Each reward location had two feeder cups that contained one 40 mg food pellet each, so that the overall amount of reward gathered per trial in this experiment was comparable to that of the other experiments. On a typical trial, the experimenter would remove the door from the start box and the rat was allowed to visit only one reward location. Rats were allowed 1 min to leave the box and visit a reward location. If a rat timed out, the trial was repeated at the end of the sequence. Testing ended if a rat persistently timed out, and was restarted the next day. In a given day, rats were tested on eight angles five times each, randomly presented so that no angle was repeated in two consecutive trials. Seven rats were run on this experiment until they successfully completed three full days.

Distance versus reward experiment. The layout was similar to the orientation experiment above, except that the starting orientation was always held at 0°, and the reward locations on each side could be at either near or far positions (figure 6).

The near positions were the same locations used in the orientation experiment. The far positions were in the same ±30° directions, but were located 110 cm from the starting location. There were four possible categories of configuration (figure 6). In the first, both targets were at the same distance from the start box quantities, but had different reward with one side having three baited cups while the other had only one. In the second category, both sides had the same reward quantity (two cups), but were located at different distances. In the third, one side was closer and had greater reward. Finally, in the fourth category one side had more reward while the other was closer. Trials were counterbalanced with respect to left, right, near and far. Rats were tested on all possible configurations within each category each day. Each of these tests consisted of four trials. The first trial was an exploratory trial in which the rat was allowed to visit and learn the locations of both goal locations. In the subsequent three trials, rats were only allowed to visit one location. Performance criteria were the same as for the orientation experiment, and all rats were run until they completed three full days. Seven rats were included in this group.

N–1 experiment. Four rats were allowed ten trials to learn a five-city configuration. On the 11th trial, one of the locations was unbaited, with the cup left in place, and rats were given up to 25 additional trials to exclude the unbaited cup from their route (figure 7). The experiment ended when the rat excluded the cup three consecutive times, or after completing the 25th additional trial. We chose to unbait the cup furthest from other cups in each configuration. The unbaited cup was never the first or last cup visited in the route the rat settled on by the end of the tenth trial.
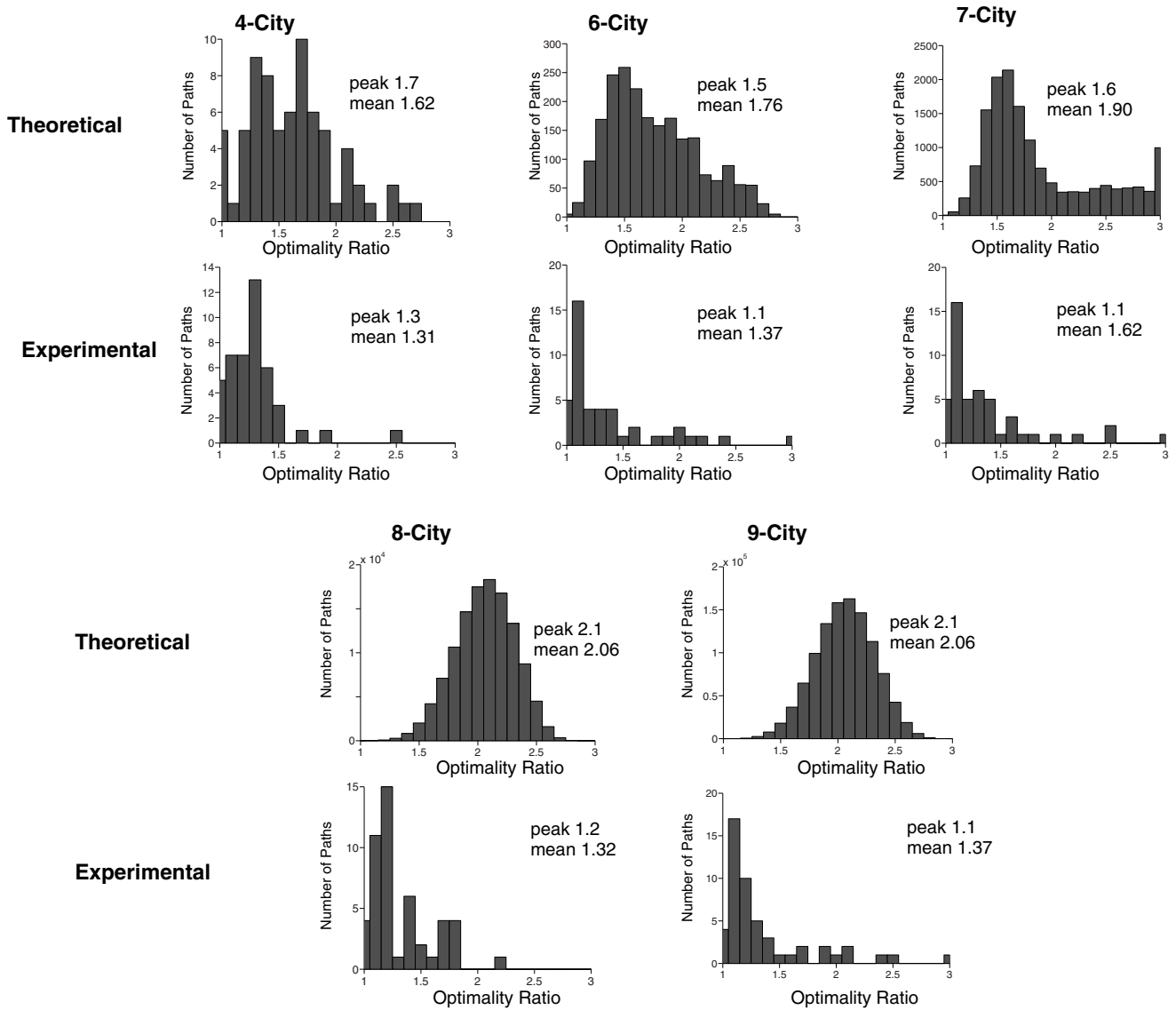
Figure 3. Theoretical and experimental pathlength distributions of the last three trials for four–nine cities. The five-city case is shown in figure 1(D).
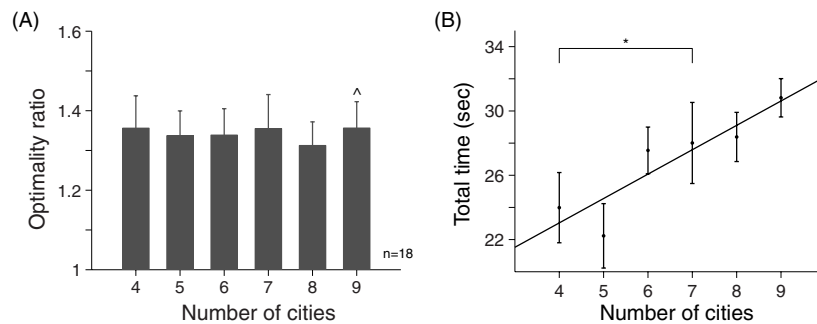
Figure 4. Rats find short route solutions to the 'variable-*N*' TSP. (A) Average optimality ratio for the last three trials for configurations with four–nine cities. The *n* corresponds to the number of configurations. (B) Average time per trial spent solving the TSP during the last three trials as a function of the city number. The number of city configurations tested were 16, 12, 18, 18, 19, 20 for four, five, six, seven, eight and nine cities, respectively.
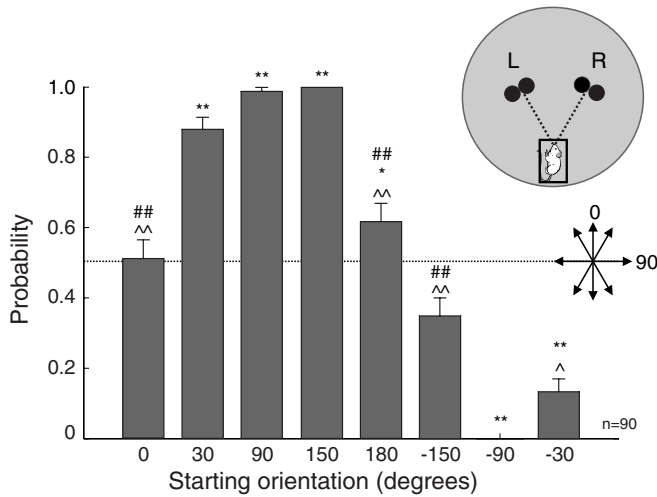
Data analyses. Rats' instantaneous positions were tracked by an overhead camera and imported into MATLAB. A custom script determined the most optimal straight-line path for each configuration using an exhaustive search method. The

optimality ratio for each trial was calculated as

Optimality ratio

$$= (\text{Length of path taken})/(\text{Length of optimal path}),$$

Figure 5. Influence of orientation on path choice. Probability of choosing the city on the right in an ambiguous configuration as a function of the starting orientation of the animal (inset). Each starting orientation was presented a total of 90 times. All tests are repeated measure ANOVAs between the data and ten randomly generated distributions of choices. * = within subjects, # = between subjects, ˆ = interaction term.

where the length of path taken was found by integrating the actual trajectory of the rat numerically in MATLAB.

Since the reflective jacket was located behind the forepaw of the rat, the measured path could be slightly shorter than the actual distance between reward locations. *Post hoc* analyses revealed that this distance discrepancy was about

10%, and all distances were corrected accordingly. In all statistical analyses (univariate repeated measures analysis), ## = p(between subjects) <0.01, # = p(between subjects) <0.05, ** = p(within subjects) <0.01, * = p(within subjects) <0.05, ˆˆ = p(interaction) <0.01, ˆ = p(interaction) <0.05. In figures 5 and 6, each dataset (bar) was tested to assess its significance with respect to random choices. For each dataset, ten independent distributions of random choices were generated, and the repeated measure ANOVA was conducted on all ten pairs (data, random-distribution-n) and averaged. For all figures, error bars are the standard error to the mean.

Results

Rats find short route solutions

Effective route planning requires the ability to spatially relate multiple goal locations within the broader context of the environment. Since obtaining this information in real space requires learning via repeated exposure to multiple place sequences in the environment, we reasoned that rats would choose shorter routes as they grew more familiar with a configuration. We tested this hypothesis in a five-city configuration (figure 1(A)). Figure 1(B) plots the optimality ratio obtained at each of ten trials averaged across seven rats and across the 24 five-cup configurations presented. Trial 1 corresponds to the first exposure of the rat to the maze. In this graph, only the trials in which the animals did not revisit a cup were considered. The average optimality ratio decreases
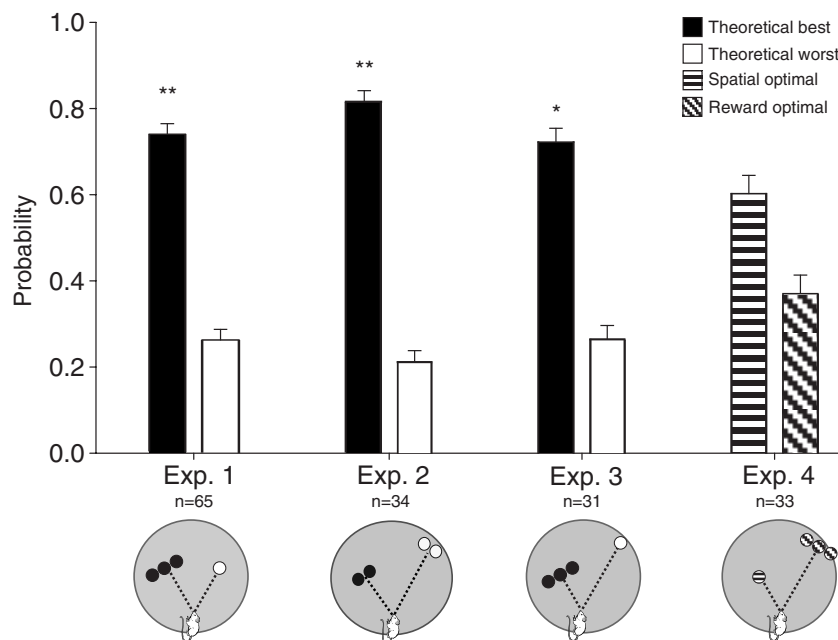
Figure 6. Relative influence of distance and reward on path choice. Four experiments (Exp. 1–4) were conducted and are schematized under the *X*-axis. For Exp. 1–3, the solid black bars correspond to the probability that rats chose the theoretically better option, while the white bars correspond the probability that the rats chose the theoretically worse option. For Exp. 4, the horizontally striped bar corresponds to the probability that rats chose the closer reward (spatially optimal), while the diagonally striped bar corresponds to the probability that the rats chose the larger reward (reward optimal). The '*n*' value corresponds to the number of times the configuration was presented. Exp. 1 has a higher '*n*' because it was counterbalanced with respect to near and far positions. Statistical significance assesses the extent to which the choices were different from random.
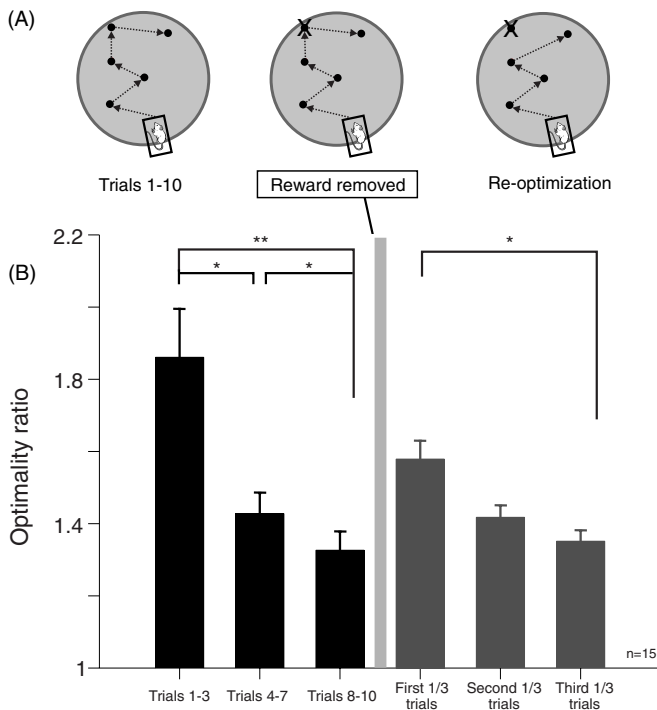
Figure 7. Dynamic re-routing. (A) Cartoon view of the experiment. (B) Optimality ratios for trials before (black) and after (gray) reward removal. Since rats varied in the number of trials to reoptimize, post-removal trials are grouped into thirds. A total of 15 configurations were tested.

linearly ($R = 0.67$, slope $= -0.013$, offset $= 1.32$) across trials, becoming significantly statistically different from trial 1 by the third trial ($n = 91$, $p < 0.05$). Trials 1 and 10 are significantly different ($n = 88$, $p < 0.01$). Figure 1(C) shows the average number of revisits per trial. This quantity decreases linearly as training progresses ($R = 0.66$, slope $= -0.06$, offset $= 1.29$). Rats typically revisited one or two cities on the first trials, and quickly settled into a five-city, no revisit configuration by the last three of the ten trials. These initial revisits can be explained by the fact that unlike with the traditional TSP where all cities are immediately accessible to the human or computer algorithm, in this rodent paradigm food cups have to be 'discovered' through active exploration during the first two–three trials. In principle, the reduction in optimality ratio can be due to a combination of three factors: (1) a better choice of city order (the core of the 'classic' TSP problem, figure 1(B)); (2) an improvement in the ability to not revisit cities (figure 1(C)); and (3) an improvement in trajectory planning between two cities (rats adopt a straighter line/trajectory between any two cities from trial to trial). Figure 2(A) shows a sample set of trajectories between two cities for different trials and suggests that rats did not significantly improve their trajectories across trials, adopting a near linear route. Figure 2(B) shows the data pooled across multiple rats and configurations. No differences are seen between the first and last three trials. Altogether, these results suggest therefore that the improvement in performance seen across trials is not significantly due to rats adopting a straighter path between pairs of cities.

We then compared the optimality ratios of the routes chosen by rats to the optimality ratios of all possible routes. In order to visit each one of five cities with no revisits, five! different paths may be taken. Figure 1(D) top shows the theoretical distribution of optimality ratios of all possible routes for the 24 five-city configurations used in our experiments. The distribution is unimodal with a peak at 1.5 and a mean at 1.57, with 25% of the routes falling below a ratio of 1.37. Figure 1(D) bottom shows the experimental distribution of optimality ratios realized by seven rats in the same 24 five-city configurations. The distribution is significantly skewed toward low ratios (peak at 1.0, mean at 1.34) with 71.4% of the paths residing below 1.37. Figure 3 shows the same theoretical and experimental distributions for four, six, seven, eight and nine cups. In all cases, the experimental distribution is strongly skewed toward near-1 optimality ratios. This result suggests that, rather than choosing routes randomly, rats naturally and quickly use spatial navigation strategies that emphasize short routes, compatible with previous studies where rats had been trained over many days (Bures *et al* 1992). Random route selection is of course the worst possible way to navigate in this task. Further work should be conducted to determine the 'default' strategies used by naïve rats (e.g. on trials 1–3), and if these default strategies exist, determine the extent to which they vary between individual rats.

One of the major challenges presented by strategies and algorithms designed to solve the TSP is to minimize the increase in the number of computational steps required to solve configurations with increasing number N of city locations. To date, the complexity of the best non-heuristically based computer algorithm increases at a rate of 2^N (Held and Karp 1962, Karp 1982). To begin to assess whether and how rats were able to find short routes in the 'variable-N' version of the TSP, in a separate set of experiments we exposed the animals to configurations consisting of four–nine rewarded locations. Figure 4(A) shows the average optimality ratios during the last three trials for each of these configurations. For each level of complexity, three different configurations were presented. We found no trend toward either improved or degraded performance as a function of the number of cities, suggesting that rats were able to find short routes within ten trials whether the number of cities was four, five, six, seven, eight or nine. Interestingly, our datasets showed a slightly lower performance for eight cities than for seven and nine, possibly indicating a change of strategy by the animals. The difference was not statistically significant, so it was not pursued further. Figure 4(B) shows the total time spent by the animals in the task during the last three trials. The data points are well described by a linear fit ($R = 0.81$, slope $= 1.52$, offset $= 21.52$). This linear increase in time is most likely attributable to the greater distance required to visit more locations, and the additional time taken to eat the food, rather than to an increase in decision-making time. Together, these results suggest that, within the limitations of our experimental paradigm, as the complexity of the task increases rats find a short route solution within ten trials and that the time spent completing the task increases linearly with N.

Influence of orientation on route selection

Due to the fact that computer agents do not typically have physical bodies, the TSP is traditionally simply viewed as a minimization of linear travel distance. Since rats have physical bodies that become directionally oriented in space as a function of their past trajectory, they may simultaneously minimize both linear travel distance and angular integration, two relatively independent constraints. This hypothesis is substantiated by the existence of populations of head-direction cells that carry orientation-specific information into the spatial navigation system (Taube 2010).

We examined this possibility by presenting rats with a configuration featuring two equidistant cups containing the same quantity of rewards and by systematically and randomly varying their starting orientation (figure 5, inset). Figure 5 shows the probability of choosing the reward location on the right (marked *R*) as a function of starting orientation (inset). A zero angle represents a starting orientation directly in between the two cups. For this starting orientation, the rats chose the right and left goals with equal probability. Individual rats, however, tend to settle on one side or the other. For $30°$, $90°$ and $150°$, rats were more likely to choose the location on the right, while for $-30°$ and $-90°$, they were more likely to choose the location on the left (marked L). These results are consistent with minimizing the angular integration to arrive at a reward location. Interestingly, for $180°$ and $-150°$, rats were more likely to choose the goal location on the right. This suggests that when visual information about reward location is poor, as when the animal faces away from goals, a left-turning bias emerges; but when visual input is available, as in the forward-facing orientations, rats are able to use it to minimize angular integration. This result is compatible with earlier work and may be attributable to asymmetries of the hippocampus (Glick 1985, LaMendola and Bever 1997, Xiao and Jordan 2002, Lister *et al* 2006).

Relative influence of reward quantity and spatial distance on route choice

Route selection in real space is likely a highly dynamical decision-making process that takes into account multiple sensory and non-sensory modalities. As expressed in the majority of computational models, these streams of information can be thought of as components of a general 'cost function' that needs to be minimized (Floresco *et al* 2008). Whether this cost function is explicitly computed in the brain, or emerges implicitly as a result of self-organizing computational principles, is unknown. Others have studied effort and reward-seeking activity using various instrumental tasks, including lever presses with delayed or probabilistic rewards and barrier climbing (Roesch *et al* 2007, Bardgett *et al* 2009, Salamone *et al* 2009, Wanat *et al* 2010). We study the extent to which route selection in this task might rely on spatial and/or reward information. To determine how rats differentially weight distance and reward cues, we held the starting orientation as constant while varying reward amount and location independently of one another in a two-city task with four different conditions (figure 6). Figure 6 shows

the probability of choosing each of the two goals for each experiment on the third trial after an initial exploration trial. In experiment 1, where the reward amount is the determining factor, rats chose the goal with the largest reward (black bars versus white bars respectively, $n = 65$, p (within) < 0.01). In experiment 2, where space is the determining factor, rats chose the closest of the two goals ($n = 34$, p (within) < 0.01). In experiment 3, where the large reward is closest and the small reward is further away, rats chose the goal with spatial and reward optimality significantly more often than the non-optimal solution ($n = 31$, p (within < 0.05). Interestingly, although the solution is optimal for both reward amount and distance, the average probability of it being chosen is no different than in experiments 1 and 2, where only one of these two features is disadvantageous. Finally, in experiment 4, where the animal is presented with an ambiguous situation in which one choice is advantageous because of its distance, and the other is advantageous because of its reward quantity, rats chose either goals with near-equal probability, with a trend toward preferring spatially optimal solutions (horizontal hatched bar, $n = 33$). This trend may change with a larger discrepancy between reward and spatial distance differences, but this study was not conducted here. It should also be noted that even though experiments 1–4 were ordered randomly, overall, and in the course of several days across all four experiments, rats were presented with more spatially advantageous choices than reward-advantageous choices. We cannot at this point exclude the possibility that part of the spatial bias in experiment 4 was due to this preponderance of spatially correct choices in experiments 1–3. These results demonstrate that rats may not rely solely on the nearest-neighbor strategy. Whenever there is a clearly correct choice, be it reward- or spatially advantageous, rats will optimize based on the relevant variable. However, when the cues are in conflict their decision-making strategy is more variable. In the conditions of these experiments, the contributions of the spatial and reward information seem to be approximately equal.

Dynamic path planning

If path length minimization is indeed a basic process of the rat brain, it is likely to be dynamical and affected by recent events. Specifically, rats must accurately learn to ignore locations that no longer provide rewards. To test this hypothesis, we ran animals in configurations in which, after learning, a reward was removed from one of the cities (figure 7(A)). Figure 7(B) shows the optimality ratios for the trials before and after reward removal. Since individual rats varied in the number of trials to exclude the unbaited location, the trials after reward removal are grouped into thirds. Optimality ratios were significantly lower for trials 8–10 than for trials 1–3 ($n = 15$, $p < 0.01$). This shows that rats had learned the initial configuration. After reward removal, the optimality ratio was higher for the first third of trials than for trials 8–10 because the animals tended to persistently visit the unbaited location. Note, however, that the ratio was not as high as when the configuration was first presented (trials 1–3), due to the similarity of the two

conditions. The optimality ratio for the last third of trials was significantly lower than for the first third ($n = 15$, $p < 0.05$). The middle third of the trials was not different from either of the other two thirds after the reward was removed. These results demonstrate that the extinction of a reward prediction signal for a previously learned location is sufficient to induce a dynamic recomputation of the path taken.

Discussion

We used a TSP-like paradigm in which animals were required to physically explore an open space to retrieve food (Menzel 1973, Bures *et al* 1992, Cramer and Gallistel 1997, Noser and Byrne 2010). The spatial complexity and difficulty of this task is unlike other approaches in which the subjects are asked to solve a computerized perceptual version of the TSP problem with little effort and few constraints (MacGregor *et al* 2004, Vickers *et al* 2006, Gibson *et al* 2007, Tenbrink and Wiener 2009, Desrochers *et al* 2010, Miyata and Fujita 2010). However, studies have shown that human perceptual TSP may involve spatial navigation schemes similar to those utilized in the real space TSP, and may recruit brain structures similar to those used by rodents (Wolbers *et al* 2007, Tenbrink and Seifert 2011). This analogy points to the potential importance of the spatial navigation system in solving the TSP, be it in real or virtual space.

Effective route planning relies in part on the ability to spatially relate multiple locations to each other and to the environment. Inherent to this requirement is the ability to remember past locations, evaluate the current position and plan for future paths. This complexity was emulated in our task by allowing rats to navigate freely through cities that were randomly placed in a relatively unconstrained open arena, unlike in previous studies with rats (Bures *et al* 1992). In the fixed-N and variable-N experiments we showed that rats were able to learn to find short routes through up to nine reward locations, well within ten trials. In those experiments, each spatial configuration was only experienced once, unlike in other studies where they were repeated over several days (Bures *et al* 1992, Noser and Byrne 2010). Furthermore, rats minimized their path length within each configuration and their performance on each first trial did not improve across days (data not shown). This suggests that the rats were unlikely to use long-term spatial memory, but rather were adhering to an online strategy that allowed them to dynamically plan and replan as the task or the environment required. Notably, the animals we used in our studies were not explicitly required to find short routes. No experimental constraints were placed on them to minimize time or distances or to maximize rewards. This suggests that this dynamical process is likely to be deeply embedded in their spatial navigation system. The intrinsic ability to minimize travel distance might be present in other systems as well, such as in the saccade circuit, in the early stages of the visual system (Desrochers *et al* 2010).

Another difference between our study and others is that rewards were presented at each of the cities, rather than delivered after all cities were visited (Bures *et al* 1992, Gibson *et al* 2007, Desrochers *et al* 2010). By baiting all locations,

reward contingencies were easily manipulated and controlled. This allowed us to directly address how reward information within a spatial environment affected the final path settled on by the rat. In the N-1 task, removal of the reward from a specific learned location resulted in replanning and in the exclusion of the unbaited location. This reinforces the notion that reward information dynamically modulates the spatial decision-making process, compatible with other studies (Roesch *et al* 2007, Bardgett *et al* 2009, Salamone *et al* 2009, Wanat *et al* 2010). In the classical TSP, spatial distance is traditionally the quantity that is optimized. In our experiment, however, we cannot exclude the possibility that rats were minimizing time instead. Since it is notoriously difficult to manipulate rat velocity through training, new experiments would have to be designed to test the relative contribution of time and space in this task. Also, the work described here focuses on pseudo-random city configurations. As such, no strong claims can yet be made on the specific strategies used by the animals, or on the possibility that these strategies might be different for small (e.g. four) and large (e.g. nine) numbers of cities. It will be interesting in the future to design configurations aimed at testing specific hypotheses on rat navigation strategies.

In contrast to most computer algorithms that have been designed to solve the TSP, we show here that when sensory inputs about the goals are available, rats are able to minimize the angular integration required to reach them. However, in the absence of such inputs, a left-turning bias emerged. The turning bias could be attributed at least in part to a change in reference frame of the head direction system occurring after the animal acquired visual inputs representing the reward locations. This idea is compatible with other experiments performed on a linear track, which showed that when a rat first left a starting box, place cells were initially bound to the box as a reference point but then became bound to visual cues around the reward location (Gothard *et al* 1996a, 1996b). It is interesting to note that a turning bias may also provide the animal with a default strategy, which would allow for a 'spiral' foraging exploration that would be far more efficient at covering space than a non-biased random walk. The bias could then be neutralized through learning, as distance and reward information become more readily available. Altogether, these results suggest that in addition to considering reward, time and space, the rat may also be minimizing rotation angle. To our knowledge, no computer algorithms have simultaneously attempted to minimize these quantities in this task (but see, Aggarwal *et al* (2000)).

Artificial neural network models have been used to solve the TSP in the past (Hopfield and Tank 1985, 1986). This pioneering work was, however, purely theoretical and solving the TSP was used as an illustration for what such networks could do, with no consideration for how animals and humans actually performed the task. Since then, tremendous progress has been made in understanding the neural substrate of spatial navigation, planning and reward processing. Correspondingly, neurally inspired computational models have been produced to gain insights into each of these computations independently.

A promising set of models uses attractor dynamics to explain the large body of experimental data gathered from the hippocampus and entorhinal cortex (Fuhs and Touretzky 2006, Navratilova *et al* 2011). These models rely on a map-like neural representation of the environment to update instantaneous positions using velocity and orientation information. Their dynamical nature seems appropriate as a starting point for modeling path planning problems such as the TSP. Other models have a more static view and propose the existence of specialized neurons that are designed to specifically encode 'short-cuts' between goal locations. These short-cuts are stored into a synaptic matrix that can be queried when necessary (Kubie and Fenton 2009). Finally, other models use a purely probabilistic yet dynamical approach (Thrun *et al* 2005). Models of this type have been implemented in robots in situations where self-motion signals and visual cues provide uncertain information about location (Milford *et al* 2010). These robot models use Bayesian approaches to optimize sensory information in order to get an accurate estimate of the current spatial location. It is not clear, however, how these neurally inspired models could be used to implement efficient spatial navigation in terms of distance and in terms of outcomes. Our data give insights into how neurally inspired models could be extended so that orientations, distances and values of reward locations are dynamically updated along a trajectory. Further pharmacological and electrophysiological studies in brain areas such as the hippocampus (space), the VTA (reward) and the prefrontal cortex (planning) in this task would further constrain the model and help capture the nature of the synergistic computations performed by these structures.

Acknowledgments

References

Aggarwal A, Coppersmith D, Khanna S, Motwani R and Schieber B 2000 The angular-metric traveling salesman problem *SIAM J. Comput.* **29** 697–711

Ainge J A, Tamosiunaite M, Woergoetter F and Dudchenko P A 2007 Hippocampal CA1 place cells encode intended destination on a maze with multiple choice points *J. Neurosci.* **27** 9769–79

Andersen P, Morris R, Amaral D, Bliss T and O'Keefe J ed 2007 *The Hippocampus Book* (Oxford: Oxford University Press)

Applegate D L 2006 *The Traveling Salesman Problem: a Computational Study* (Princeton, NJ: Princeton University Press)

Bagchi T P, Gupta J N D and Sriskandarajah C 2006 A review of TSP based approaches for flowshop scheduling *Eur. J. Oper. Res.* **169** 816–54

Bardgett M E, Depenbrock M, Downs N, Points M and Green L 2009 Dopamine modulates effort-based decision making in rats *Behav. Neurosci.* **123** 242–51

Blum A, Chawla S, Karger D R, Lane T, Meyerson A and Minkoff M 2007 Approximation algorithms for orienteering and discounted-reward TSP *SIAM J. Comput.* **37** 653–70

Bures J, Buresova O and Nerad L 1992 Can rats solve a simple version of the traveling salesman problem? *Behav. Brain Res.* **52** 133–42

Cardema J C, Wang P K C and Rodriguez G 2004 Optimal path planning of mobile robots for sample collection *J. Robot. Syst.* **21** 559–80

Cramer A E and Gallistel C R 1997 Vervet monkeys as travelling salesmen *Nature* **387** 464

Desrochers T M, Jin D Z, Goodman N D and Graybiel A M 2010 Optimal habits can develop spontaneously through sensitivity to local cost *Proc. Natl Acad. Sci. USA* **107** 20512–7

Dupret D, O'Neill J, Pleydell-Bouverie B and Csicsvari J 2010 The reorganization and reactivation of hippocampal maps predict spatial memory performance *Nat. Neurosci.* **13** 995–1002

Floresco S B, St Onge J R, Ghods-Sharifi S and Winstanley C A 2008 Cortico-limbic-striatal circuits subserving different forms of cost-benefit decision making *Cogn. Affect. Behav. Neurosci.* **8** 375–89

Fuhs M C and Touretzky D S 2006 A spin glass model of path integration in rat medial entorhinal cortex *J. Neurosci.* **26** 4266–76

Gallistel C R and Cramer A E 1996 Computations on metric maps in mammals: getting oriented and choosing a multi-destination route *J. Exp. Biol.* **199** 211–7

Gibson B M, Wasserman E A and Kamil A C 2007 Pigeons and people select efficient routes when solving a one-way 'traveling salesperson' task *J. Exp. Psychol. Anim. Behav. Process* **33** 244–61

Glick S D 1985 Heritable differences in turning behavior of rats *Life Sci.* **36** 499–503

Gothard K M, Skaggs W E and McNaughton B L 1996a Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues *J. Neurosci.* **16** 8027–40

Gothard K M, Skaggs W E, Moore K M and McNaughton B L 1996b Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task *J. Neurosci.* **16** 823–35

Gutin G and Punnen A P 2002 *The Traveling Salesman Problem and its Variations* (Dordrecht: Kluwer)

Held M and Karp R M 1962 A dynamic programming approach to sequencing problems *J. Soc. Indust. Appl. Math.* **10** 196–210

Hok V, Lenck-Santini P P, Roux S, Save E, Muller R U and Poucet B 2007 Goal-related activity in hippocampal place cells *J. Neurosci.* **27** 472–82

Hopfield J J and Tank D W 1985 'Neural' computation of decisions in optimization problems *Biol. Cybern.* **52** 141–52

Hopfield J J and Tank D W 1986 Computing with neural circuits: a model *Science* **233** 625–33

Karp R M 1982 Dynamic programming meets the principle of inclusion and exclusion *Oper. Res. Lett.* **1** 49–51

Kubie J L and Fenton A A 2009 Heading-vector navigation based on head-direction cells and path integration *Hippocampus* **19** 456–79

LaMendola N P and Bever T G 1997 Peripheral and cerebral asymmetries in the rat *Science* **278** 483–6

Lister J P, Tonkiss J, Blatt G J, Kemper T L, DeBassio W A, Galler J R and Rosene D L 2006 Asymmetry of neuron numbers in the hippocampal formation of prenatally malnourished and normally nourished rats: a stereological investigation *Hippocampus* **16** 946–58

MacGregor J N, Chronicle E P and Ormerod T C 2004 Convex hull or crossing avoidance? Solution heuristics in the traveling salesperson problem *Mem. Cognit.* **32** 260–70

Marshall J A, Bogacz R, Dornhaus A, Planque R, Kovacs T and Franks N R 2009 On optimal decision-making in brains and social insect colonies *J. R. Soc. Interface* **6** 1065–74

Martig A K and Mizumori S J 2011 Ventral tegmental area disruption selectively affects CA1/CA2 but not CA3 place

fields during a differential reward working memory task *Hippocampus* **21** 172–84

Menzel E W 1973 Chimpanzee spatial memory organization *Science* **182** 943–5

Milford M J, Wiles J and Wyeth G F 2010 Solving navigational uncertainty using grid cells on robots *PLoS Comput. Biol.* **6** e1000995

Miyata H and Fujita K 2010 Route selection by pigeons (*Columba livia*) in 'traveling salesperson' navigation tasks presented on an LCD screen *J. Comp. Psychol.* **124** 433–46

Mizumori S J 2008 *Hippocampal Place Fields: Relevance to Learning and Memory* (Oxford: Oxford University Press)

Navratilova Z, Giocomo L M, Fellous J M, Hasselmo M E and McNaughton B L 2011 Phase precession and variable spatial scaling in a periodic attractor map model of medial entorhinal grid cells with realistic after-spike dynamics *Hippocampus* at press

Noser R and Byrne R W 2010 How do wild baboons (*Papio ursinus*) plan their routes? Travel among multiple high-quality food sources with inter-group competition *Anim. Cogn.* **13** 145–55

Roesch M R, Calu D J and Schoenbaum G 2007 Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards *Nat. Neurosci.* **10** 1615–24

Salamone J D, Correa M, Farrar A M, Nunes E J and Pardo M 2009 Dopamine, behavioral economics, and effort *Front. Behav. Neurosci.* **3** 13

Schultz W 2010 Dopamine signals for reward value and risk: basic and recent data *Behav. Brain Funct.* **6** 24

Taube J S 2010 Head direction cell firing properties and behavioral performance in 3D space *J. Physiol* **589** 835–41

Tenbrink T and Seifert I 2011 Conceptual layers and strategies in tour planning *Cogn. Process.* **12** 109–25

Tenbrink T and Wiener J 2009 The verbalization of multiple strategies in a variant of the traveling salesperson problem *Cogn. Process.* **10** 143–61

Thrun S, Burgard W and Fox D 2005 *Probabilistic Robotics* (Cambridge, MA: MIT Press)

Vickers D, Lee M D, Dry M, Hughes P and McMahon J A 2006 The aesthetic appeal of minimal structures: judging the attractiveness of solutions to traveling salesperson problems *Percept. Psychophys.* **68** 32–42

Wanat M J, Kuhnen C M and Phillips P E 2010 Delays conferred by escalating costs modulate dopamine release to rewards but not their predictors *J. Neurosci.* **30** 12020–7

Wolbers T, Wiener J M, Mallot H A and Buchel C 2007 Differential recruitment of the hippocampus, medial prefrontal cortex, and the human motion complex during path integration in humans *J. Neurosci.* **27** 9408–16

Xiao L and Jordan C L 2002 Sex differences, laterality, and hormonal regulation of androgen receptor immunoreactivity in rat hippocampus *Horm. Behav.* **42** 327–36