

# Optimizing and “Pessimizing”: Human Performance with Instructional Variants of the Traveling Salesperson Problem

*Edward Chronicle,<sup>1</sup> James MacGregor,<sup>2</sup> and Thomas Ormerod<sup>3</sup>*

## Abstract

The two-dimensional Traveling Salesperson Problem (TSP) requires finding the shortest tour through  $n$  locations. Untrained adults are adept at the task and reliably outperform simple construction algorithms for  $n \leq 60$ . Performance may stem from a specific inherent ability. Alternatively, it may reflect general spatial intelligence, whether inherent or acquired. If the latter holds, then people should be equally adept at finding longest tours. Two experiments comparing ability in the two tasks found participants significantly better at finding short tours than long tours (in the first, with  $N = 30$ ,  $F[1,28] = 14.14$ ;  $\eta_p^2 = .34$ ;  $p < 0.01$ , and in the second, with  $N = 112$ ,  $F[1,108] = 48.76$ ;  $\eta_p^2 = .31$ ;  $p < 0.01$ ). Furthermore, human performance was significantly worse than a simple construction algorithm (farthest-neighbor) for the task of finding long tours. The result is consistent with the hypothesis of a specific, inherent ability to find short routes.

## Introduction

The two-dimensional Euclidean Traveling Salesperson Problem (TSP) is an optimization problem, the standard version of which is to find the shortest closed tour through a set of locations in the plane (Figure 1). TSPs are common in industrial and management settings and can also be recognized in natural behaviors, such as minimizing energy expended in visiting locations of potential food sources. They present a challenge because of their combinatorial properties: The number of tours is given by  $(n-1)!/2$ , where  $n$  is the number of locations. As  $n$  increases it becomes unfeasible to sift through all possible tours to find the best. Rather, algorithms have been developed to produce acceptably short tours at low computational cost. While none can guarantee to find the best tour for any problem in reasonable time, it is now the case that sophisticated algorithms running on computer networks have found optimal tours for very large problem instances, such as the 24,978 cities of Sweden problem (see [www.tsp.gatech.edu/index.html](http://www.tsp.gatech.edu/index.html)).

---

<sup>1</sup>University of Hawaii at Manoa

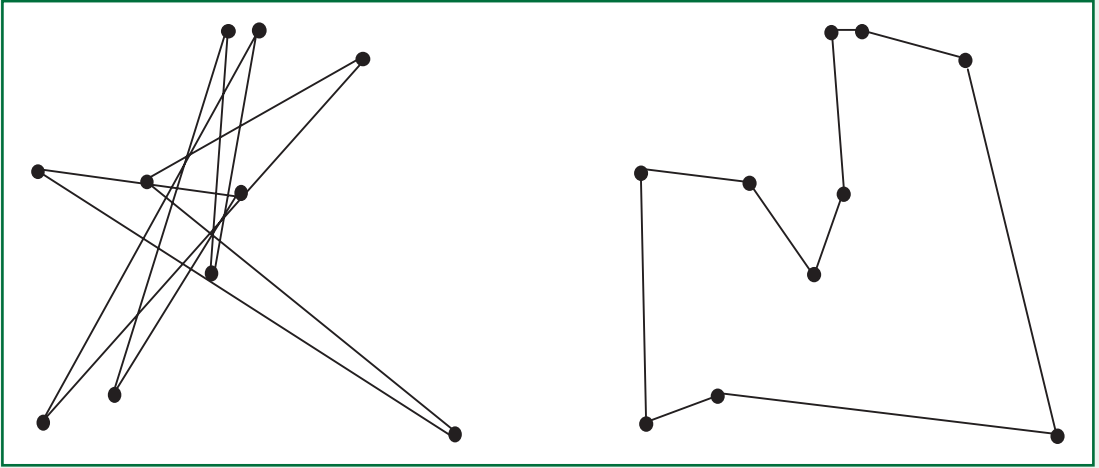
<sup>2</sup>University of Victoria, B.C., Canada

<sup>3</sup>Lancaster University, UK

Corresponding author: Edward P. Chronicle, Ph.D., Department of Psychology, University of Hawaii at Manoa, 2430 Campus Road, Honolulu, HI 96822

**Figure 1.**

*Longest and shortest tours for a TSP with  $n = 10$ .*



We have previously demonstrated that untrained adults reliably outperform simple construction algorithms in problems where  $n \leq 60$  (MacGregor & Ormerod, 1996). We proposed that this surprising ability may be inherent, an interpretation that is consistent with findings that other species also “solve” TSPs (Cramer & Gallistel, 1997; Menzel, 1973). Indeed, it is easy to imagine scenarios—such as foraging for food—where such a capacity has an evolutionary advantage. Conversely, this ability may not be inherent. It may simply reflect general spatial intelligence and specifically an ability to judge proximity relations (Vickers, Mayo, Heitmann, Lee, & Hughes, 2004).

A variant of the standard problem is to find the longest tour (an example is given in Figure 1). For Euclidean distances in three or more dimensions this version is as complex as the classic TSP, and for two dimensions, it is conjectured to be so (Barvinok, Fekete, Johnson, Tamir, Woeginger, & Woodroffe, 2003). For the remainder of this paper, we will use the following terms to refer to TSP tours. The *shortest tour* is the known solution to the standard version of the TSP (often referred to as the optimal tour). The *longest tour* is the known solution to the variant version. *Short tours* are those tours produced by human participants attempting to generate the shortest tour, and *long tours* are those produced by human participants attempting to generate the longest tour. *Nearest-neighbor* and *farthest-neighbor* tours are those produced by simple serial construction algorithms that, respectively, select the nearest and farthest location as the next one in a tour sequence.

On prima facie grounds, it is less easy than for the standard version of the TSP to imagine scenarios where finding the longest tour would have any evolutionary value. If the ability to find short tours is inherent, then there is no reason to expect that humans will be equally proficient at finding long tours. Conversely, if finding short tours reflects general spatial intelligence, one might expect performance in finding maximum and

minimum tours to be equally impressive. Such considerations motivated us to compare the tasks experimentally.

## Experiment 1

### *Method*

#### *Participants*

Twenty volunteers were recruited from the campus community at Lancaster University, UK. Identifiers were not collected, so age and gender were not known. All participants were naive to the TSP and had not received any information about human optimization problems during their classes. Ethical-committee review of this experiment was not required in the UK.

#### *Materials*

Five 10-point TSPs were generated randomly, and each was printed on U.S. letter paper within a coordinate space measuring approximately  $10 \times 10$  cm, with points marked by black circular dots 2.5 mm in diameter. Participants received booklets of problems and were simply required to construct their tours using a pen or pencil.

#### *Design and Procedure*

A single variable, tour type (the requirement to produce long or short tours), was manipulated within subjects. Each participant produced long and short tours to the same five TSPs. Tour type was blocked and counterbalanced across participants, and the order of TSPs was randomized within each block. Prior to each block of problems, appropriate instructions were provided on a separate page of the problem booklet. Before starting the experiment, participants undertook two practice problems and were given the opportunity to seek clarification from the experimenter: This was never required in practice. The experiment was conducted in an individual setting in a cognitive psychology laboratory. The time taken to construct each tour was recorded by the experimenter using a handheld stopwatch.

### *Results*

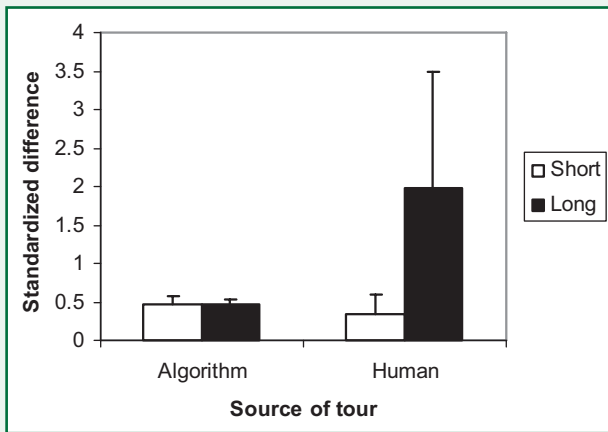
Tours from each participant were manually coded by reference to a coordinate template. Tours were excluded if (a) points had been visited more than once, (b) points had been omitted, or (c) point-to-point lines were obscured or indecipherable. Tour distances were calculated trigonometrically and tabulated. Shortest and longest tours were found with an algorithm developed independently in the second author's laboratory. These were confirmed using the Purdue TSP App ([psych.purdue.edu/tsp/workshop/downloads.html](http://psych.purdue.edu/tsp/workshop/downloads.html)), which is a downloadable application allowing the computation of TSP

tours using well-recognized algorithms. Sets of nearest-neighbor and farthest-neighbor tours were also computed using each of the ten available starting points in each problem. Each instance of a computed tour was regarded as equivalent to one human tour; as this experiment used 10-point problems, there were 10 nearest-neighbor and 10 farthest-neighbor tours available for each problem. These tours were used as comparators for human-generated data in the subsequent analysis, with means being taken across problem as appropriate.

Tour distances were then converted into standardized differences from the shortest and longest tours, using standard deviations computed from 10,000 Monte Carlo tours for each problem. The resultant standardized differences were averaged across problem and submitted to analysis of variance, with tour type (long, short) as a within-participant factor and source of tour (human, computed heuristic) as a between-participant factor. Mean standardized differences are plotted in Figure 2.

**Figure 2.**

*The interaction between source of tour and tour type in Experiment 1. Error bars show standard deviations.*



There was a significant interaction between tour type and source of tour ( $F[1,28] = 14.14; \eta_p^2 = .34; p < 0.01$ ). Human participants were much more adept at constructing short tours than long, whereas the computed nearest-neighbor and farthest-neighbor tours gave rise to almost identical standardized differences. Of note, 31 of the 100 short tours produced by human participants were minimal, whereas none of the long tours was maximal. Simple effects analysis demonstrated that participants were slightly but not significantly better than the nearest-neighbor heuristic at constructing short tours ( $F[1,56] = 0.16; p = 0.69$ ) but were significantly worse than the farthest-neighbor heuristic at constructing long tours ( $F[1,56] = 19.26; \eta_p^2 = .60; p < 0.001$ ).

Human performance on short tours was not only better than on long, it was also faster. On average short tours took significantly less time than long tours, at 5.67 s versus 12.76 s,  $t(19) = 9.67$ ; Cohen's  $d = 2.23; p < .001$ .

## *Discussion*

Experiment 1 revealed a clear dissociation in human performance between the tasks of generating short and long tours in 10-point TSPs. When generating long tours, humans performed significantly worse than when generating short tours, and also significantly worse than the computed farthest-neighbor heuristic. On average, human long tours were nearly two standard deviations shorter than the true longest tour, whereas their short tours were less than half a standard deviation longer than the true shortest tour.

Taken at face value, the results of Experiment 1 support the notion that humans may have access to an optimizing heuristic but not to a “pessimizing” heuristic. It is possible, however, that other factors contributed to the poor human performance at generating long tours. One notable feature of the long tours was that points of the problem were occasionally missed and/or lines redrawn. As [Figure 1](#) shows, long tours require multiple line crossings, and it is conceivable that participants became frustrated or confused because their ability to monitor their performance during the task may have been obscured by the visual clutter of lines already drawn. Experiment 2 was designed to overcome this potential problem by introducing a second manipulation: Participants were asked either to generate their own tours or explicitly to follow nearest- and farthest-neighbor heuristics. On the one hand, if visual clutter is the sole determinant of poor performance on long tours, we expect that generation and heuristic-following performance would be similar. On the other hand, if generation of long tours is inherently difficult for the reasons suggested above, generating them should give rise to worse performance than following a farthest-neighbor heuristic.

## Experiment 2

### *Method*

#### *Participants*

A total of 114 volunteers were recruited from introductory psychology classes at the University of Hawaii at Manoa. Two provided no data and were thus excluded from analysis. As in Experiment 1, identifiers were not collected, and participants were naive to the TSP. Class credit was awarded for participation. Ethical approval for this experiment was given by the Committee on Human Subjects of the University of Hawaii.

#### *Materials*

Six 15-point TSPs were generated randomly, with the constraint that 6 points should appear on the convex hull of each problem. Problems were printed on U.S. letter paper within a coordinate space measuring approximately  $10 \times 15$  cm, with points marked by

black circular dots, 2 mm in diameter. Participants received booklets of problems and were simply required to construct their tours using a pen or pencil.

### *Design and Procedure*

Two variables were manipulated in a factorial, between-participants design. The first variable was tour type: the requirement to produce long or short tours. The second variable was task type: Participants were instructed either to generate tours or to follow instructions for farthest- or nearest-neighbor heuristics. Each participant received a booklet containing the six TSPs in random order, prefaced by the appropriate instructions for his or her condition. The experiment was conducted in a group setting in a university classroom: A total of 15 minutes was allowed for completion of the booklet.

### *Results*

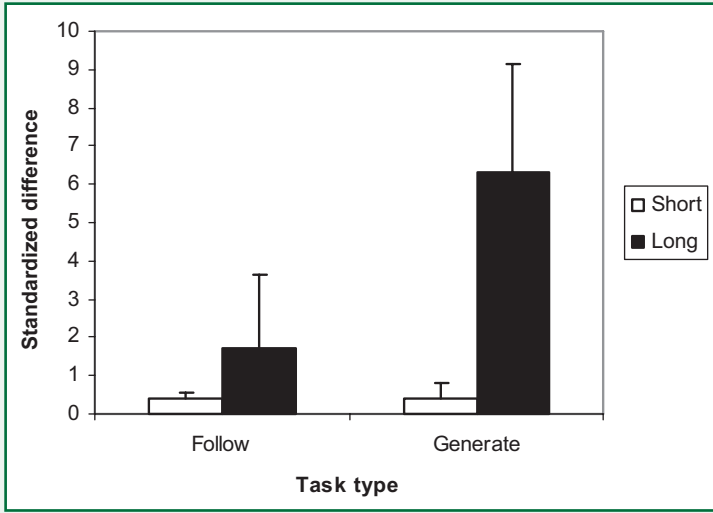
Tours from each participant were coded manually in a similar fashion to Experiment 1. Shortest and longest tours were found using the Purdue TSP App ([psych.purdue.edu/tsp/workshop/downloads.html](http://psych.purdue.edu/tsp/workshop/downloads.html)); shortest tours were confirmed by an algorithm developed independently in the second author's laboratory, but it was not possible independently to confirm longest tours. Sets of heuristic-following tours were then computed, using each of the fifteen available starting points in each problem, and the mean distance of farthest-neighbor and nearest-neighbor tours was calculated.

Tour distances from participants were then converted into standardized differences from the appropriate reference point (shortest tour, longest tour, mean farthest-neighbor tour, mean nearest-neighbor tour), using standard deviations computed from 10,000 Monte Carlo tours for each problem. The resultant standardized differences were averaged across the six TSPs and submitted to analysis of variance, with tour type (long, short) and task type (generate, follow) as between-participant factors. Mean standardized differences are plotted in [Figure 3](#).

There was a significant interaction between tour type and task type ( $F[1,108] = 48.76; \eta_p^2 = .31; p < 0.01$ ). Participants were substantially worse at generating long tours than short tours, but their farthest-neighbor following performance was much closer to their nearest-neighbor following performance (see [Figure 3](#)). Simple effects analysis confirmed (a) that generating long tours resulted in significantly poorer performance than following the farthest-neighbor heuristic ( $F[1,108] = 95.11; \eta_p^2 = .47; p < 0.001$ ) and (b) that participants are significantly better at generating short tours than long tours ( $F[1,108] = 167.44; \eta_p^2 = .61; p < 0.001$ ). In addition, participants were significantly better at following the nearest neighbor heuristic than the farthest neighbor ( $F[1,108] = 7.70; \eta_p^2 = .07; p < 0.01$ ) and were slightly but not significantly better at generating short tours than the nearest neighbor heuristic ( $F[1,108] = 2.81; \eta_p^2 = .03; p = 0.097$ ). There were

**Figure 3.**

The interaction between task type and tour type in Experiment 2. Error bars show standard deviations.



significant main effects of both tour requirements, with long tours having significantly larger mean standardized differences from reference point than short tours ( $F[1,108] = 120.48$ ;  $\eta_p^2 = .53$ ;  $p < 0.01$ ) and task type, with heuristic-following performance giving rise to significantly smaller mean standardized differences from reference point than generation performance ( $F[1,108] = 48.17$ ;  $\eta_p^2 = .31$ ;  $p < 0.01$ ).

## General Discussion

In two experiments, we investigated the ability of human participants to optimize and pessimize—that is, to find the longest possible tour—their performance on TSPs. Experiment 1 simply required participants to generate both long and short tours to 10-point problems. Across problems, short tours were significantly closer to the shortest tour than were long tours to the longest. Furthermore, human participants on average found slightly shorter tours than a nearest neighbor algorithm; however, a farthest-neighbor algorithm found significantly longer tours than did humans. Experiment 2 replicated the finding of very poor performance of humans on generating long tours and suggested that this was not simply a consequence of the visual clutter inherent in long tours. Despite similar amounts of visual clutter in both generation and heuristic-following conditions, participants were significantly worse at long tour generation. It also confirmed that human performance in generating short tours was slightly better than the average performance of a nearest-neighbor algorithm. It is evident from Experiment 2, however, that there is a residual difficulty to the farthest-neighbor following task: This was significantly more difficult for participants than the nearest-neighbor following task.

We suggest that this residual difficulty may lie in the relative difficulty of judging longer distances across the stimulus. Such judgments seem likely to be influenced both by relative eccentricity of stimuli and by noise (Waugh & Levi, 1995). It is also possible (for both tour generation and heuristic-following tasks) that participants might become confused if long lines drawn between distant points “accidentally” intersect interior points; we acknowledge that this issue of stimulus design will need to be addressed in future.

If humans’ ability to find short tours were a consequence of general spatial ability, then we might reasonably expect them to be equally good at finding long tours. It should be acknowledged, however, that while human performance in finding short tours appears to correlate well with spatial intelligence (Burns, Lee, & Vickers, 2006), there are no extant data on the relationship between spatial intelligence and pessimizing. Nonetheless, the performance dissociations reported in this paper are consistent with and lend credence to our hypothesis that humans have access to a specific optimizing heuristic for finding short routes. Similarly, specific heuristics have been proposed for other analytically complex tasks, such as catching (Gigerenzer, 2004). In planar TSPs, we have suggested (Chronicle, MacGregor, Ormerod, & Burr, 2006) that the human heuristic that facilitates short tour generation may share mechanisms with low-level boundary or contour detection processes in the visual system (Marr, 1982).

## References

- Barvinok, A., Fekete, S. P., Johnson, D. S., Tamir, A., Woeginger, G. J., and Woodroffe, R. (2003). The geometric maximum Traveling Salesman Problem. *Journal of the ACM*, *50*, 641–664.
- Burns, N., Lee, M. D., and Vickers, D. (2006). Are individual differences in performance on perceptual and cognitive optimization problems determined by general intelligence? *Journal of Problem Solving*, in press.
- Chronicle, E. P., MacGregor, J. N., Ormerod, T. C., & Burr, A. (2006). It looks easy! Heuristics for combinatorial optimization problems. *Quarterly Journal of Experimental Psychology*, in press.
- Cramer, A. E., & Gallistel, C. R. (1997). Vervet monkeys as traveling salesmen. *Nature*, *387*, 464.
- Gigerenzer, G. (2004). Striking a blow for sanity in theories of rationality. In M. Augier & J. G. March (Eds.). *Models of a man: Essays in memory of Herbert A. Simon* (pp. 389–409). Cambridge, MA: MIT Press.
- MacGregor, J. N., & Ormerod, T. C. (1996). Human performance on the Traveling Salesman Problem. *Perception and Psychophysics*, *58*, 527–539.
- Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- Menzel, E. W. (1973). Chimpanzee spatial memory organization. *Science*, *182*, 943.
- Vickers, D., Mayo, T., Heitmann, M., Lee, M.D., & Hughes, P. (2004). Intelligence and individual differences in performance on three types of visually presented optimization problems. *Personality and Individual Differences*, *36*, 1059–1071.

Waugh, S. J. & Levi, D. M. (1995). Spatial alignment across gaps: Contributions of orientation and spatial scale. *Journal of the Optical Society of America A*, 12, 2305–2317.

## Acknowledgments

The research reported in this paper benefited greatly from discussions with Ulrike Stege, Department of Computer Science, University of Victoria, British Columbia. We are grateful to Caroline Wade for data collection in Experiment 1 and to various members of the first authors' laboratory for assistance with data collection and coding.