Journal of Problem Solving

Heuristics for Comparing the Lengths of Completed E-TSP Tours: Crossings and Areas

James N. MacGregor[1]

[1] *University of Victoria*

Correspondence:
Correspondence concerning this article should be addressed to James N. MacGregor, via email at jmacgreg@uvic.ca.

Keywords:
problem solving, heuristics, optimization, area judgments

Acknowledgement:
The research was partially supported by a grant from the National Research Council for Science and Engineering (Canada).

The article reports three experiments designed to explore heuristics used in comparing the lengths of completed Euclidean Traveling Salesman Problem (E-TSP) tours. The experiments used paired comparisons in which participants judged which of two completed tours of the same point set was shorter. The first experiment manipulated two factors, the presence/absence of crossed arcs, and the relative areas of the enclosed polygons. Both factors significantly influenced judgments, with the absence of crossings and smaller areas being associated with shorter tours. The second experiment examined the effects of crossings only, and compared stimulus pairs using all possible combinations of no, one, and more than one crossing. The results showed a significant tendency for tours with one or more crossings to be judged longer than tours with none, while tours with more crossings were not judged to be longer than tours with only one. Apparently the mere presence of a crossing is sufficient to cause a tour to be judged as longer. The third experiment examined the effects of area only, and consisted of two parts. In the first part, participants judged which of two tours that differed in area was shorter. The results supported those of the first experiment, by finding that tours with smaller areas tended to be judged as shorter. In the second part of the experiment, participants judged the relative areas of each pair, to determine whether people can reliably differentiate the areas of such complex polygons. The results confirmed that they can, thereby supporting the feasibility of using differences in area as a heuristic to judge relative lengths. The results were discussed in terms of Carruthers's (2015) proposal of goal modification and the suggestion is made that applying heuristics of the type identified may represent a specific form of goal modification.

INTRODUCTION

The letter of invitation to this special issue indicated that it was partly motivated by the PhD dissertation of Sarah Carruthers (2015) which, among other things, proposed that when people are tasked with finding optimal solutions to instances of hard computational problems, the goal of the task may not be well-defined and therefore not encodable. For a goal to be well-defined, it must be possible to determine when it has been achieved (p. 41), which may not be the case in many hard optimization problems. An example is the Euclidean Traveling Salesman Problem (E-TSP), where problem solvers are provided with a set of points and invited to find the shortest tour that passes through each point and returns to the starting point. It is unlikely that, in any but the most trivial instances, problem solvers could know that an optimum tour had been achieved without external aid. Carruthers considers two reasons for this, one being that people may be unable to judge accurately which of two candidate solutions is shorter (p. 41). If people are unable to determine whether an optimal tour has been attained then they cannot be working on the given task, but "on some other, unknown, task" (p. 44).

In the case of E-TSP, if people are working on some unknown task, the fact remains that their solutions are frequently very good, and sometimes optimal (Dry, Lee, Vickers, & Hughes, 2006; Graham, Joshi, & Pizlo, 2000; MacGregor & Ormerod, 1996). The question therefore persists as to how human solvers are able to reach good, fast, solutions to problems with indeterminable goal states. In the psychology of problem solving and decision making, a long-standing answer to similar questions has been that problem solvers use heuristic procedures (Gigerenzer & Goldstein, 1996; Kahneman & Tversky, 1972; Simon, 1983; Tversky & Kahneman, 1973). However, before assuming that heuristics explain performance, Carruthers (2015) proposes that identifying the problem as encoded is important, "because the encoded problem drives performance" (p. 47). The present article explores this proposal by designing a task that meets the definition of having an indeterminable goal and showing that people, first, act as if they believe they can achieve the goal and, second, appear to use heuristics in trying to do so.

The procedures described below use the method of paired comparisons to examine how people judge which of two tours

http://dx.doi.org/10.7771/1932-6246.1193

is shorter, even when the tours are of virtually identical lengths. It is one of a few studies to investigate E-TSP performance by having people judge tours rather than generate them. One of the first studies to use a judgment paradigm invited participants to assess the gestalt quality of goodness of figure of 30 completed 10-node TSP tours that ranged in length from 0% to 45% above the optimal (Ormerod & Chronicle, 1999). The results indicated that mean goodness ratings increased linearly with decreasing distance from optimality. The findings therefore suggest that figural goodness may provide one potential heuristic mechanism for judging the quality of a completed tour. Vickers et al. (2006) reported an experiment in which participants judged the perceived goodness of 162, 25-node tours that varied in distance above optimal from 0% to 25%, and in a number of additional stimulus characteristics. The reporting of results focused primarily on individual differences which, as the authors stated, did not make for "easy generalizations" (Vickers, Lee, Dry, Hughes, & McMahon, 2006, p. 38). Nevertheless, among the variables reported as contributing to figural goodness were path length, circularity, path complexity, and convexity. The independent variables were also correlated with one another, suggesting a nexus of stimulus characteristics that people could use to judge the lengths of tours.

A third study using a judgment paradigm was reported by Dry and Fontaine (2014), which, unlike the previous studies, presented unconnected point sets rather than completed tours as stimuli. On each trial, participants were shown four point sets varying on several dimensions and were asked to select which would be easiest to solve optimally (Dry & Fontaine, 2014). The results indicated that both of the stimulus properties that were independently varied—number of potential intersections and number of nodes on the convex hull—had a significant effect on judgments. An additional post hoc analysis of stimulus properties indicated that the number of indentations, a factor related to convexity, may also have influenced judgments. This latter finding stands in contrast to results reported in MacGregor (2012) from a study using a paired-comparison procedure of completed tours. In this case a stimulus was a pair of *optimal* tours of a given point set that varied in the number of indentations. Ten different stimuli were used, and participants judged which member of a pair was shorter. The results showed no significant effect of number of indentations on length judgments. However, because differences in number of indentations were either one or two, it is possible that the manipulation was not sufficiently strong to have influenced judgments. Also, the fact that participants were comparing tours that were both optimal may have placed limits on any possible effects.

Of the studies reviewed above, all but the last identified stimulus factors that people used in judging the lengths of actual or potential TSP tours. The factors included figural goodness, circularity, and convexity, any or all of which might be used as heuristics in judging whether or not an "optimal" tour has been achieved. Of note, none of the studies included the presence of intersections as a stimulus factor, and one explicitly excluded them (Vickers et al., 2006). This is notable because a well-established finding is that human tours rarely produce crossed arcs (Graham, Joshi, & Pizlo, 2000; MacGregor & Ormerod, 1996). On average, only 6% of 1826 human-generated tours showed crossings in results summarized by van Rooij, Stege, and Schactman (2003), who proposed that in constructing tours, people aim to avoid creating crossed arcs. The results suggest that people may be sensitive to crossings, which suggests that the presence or absence of crossings may also serve as a heuristic in judging tour lengths.

The present study reports three experiments that, separately and in combination, tested whether participants used (1) the presence of crossings in E-TSP tours, and (2) the areas of the polygons enclosed by them as heuristics in judging the relative lengths of tours. For purposes of control, the studies compared pairs of tours that were of virtually the same lengths and that used exactly the same point sets. Because of this, differences in area, convexity, and circularity of pair members were perfectly correlated in the present stimuli. For convenience, reference is made to "area" in most instances, recognizing that area, convexity and circularity are indistinguishable in the present case.

EXPERIMENT 1

Experiment 1 tested the hypotheses that (1) tours with crossings are judged to be longer than tours without, and (2) tours with larger areas are judged to be longer than tours with smaller areas.

METHOD

Participants. Participants were 20 volunteers recruited from the campus community at the University of Victoria.

Stimuli. The stimuli were 60 completed instances of TSPs organized into 30 paired comparisons. In 20 of the comparisons, one member of a pair had a crossing. There were 10 comparisons each of 10-node and 20-node instances. The remaining 10 comparisons involved pairs with no crossings but different areas, five of 10-node instances, and five of 20-node instances.

The stimuli were produced in the following way. A 10-node point set was randomly generated under the constraint that exactly five points fell on the convex hull, relatively close to the expected value of 5.95 for randomly generated 10-node point sets (Philip, 2004). The set was constrained to occupy a display of 512 x 512 pixels. Then, tours were randomly generated until one was found without a crossing whose length fell within the typical range of human solutions. This represented a base tour. Next, random tours were produced until one was found with a crossing whose length was (1) shorter than the base tour, and (2) within 2% of its length. This was then repeated until a tour was found whose length was within 2% of the base tour's and with a different area. This procedure resulted in a triad of tours of virtually the same length for the same point set, consisting of 2 tours without crossings but that differed in area (the base tour and the alternate area tour) and one tour with at least one crossing (the crossing tour). The procedure was then repeated until there was a total of five such triads. Finally, the whole procedure was repeated to produce a second set of five triads, each of 20-node instances (although in this case the number of nodes on the convex hull was not constrained). For each of the 10 triads, the shortest tour's length was within 1.45% of the longest. For 9 of the 10, the crossing tour was the shortest (for one case, the alternate area tour was 0.2% longer than the crossing tour). For the base and alternate area tours, the differences between the larger and smaller areas ranged from 3% to 46% with a mean of 17%. For the 10-node stimuli, base tours ranged from 3% to 23% above the optimal solution, and for the 20-node stimuli, from 11% to 39%.

Each of the 10 stimulus triads yielded three paired comparisons, base versus crossing, alternate area versus crossing, and base versus alternate area, resulting in 20 comparisons of a crossing and non-crossing tour, and 10 of two non-crossing tours that differed in area. The 30 paired comparisons were presented as Powerpoint slides in random order. On each slide, a randomly selected member of the pair was rotated through 90 degrees to help disguise the fact that each involved the same point set. An example of a stimulus pair is shown in Figure 1 (see next page).

Procedure. Participants were tested individually. Instructions stated that the task was to indicate which member of each pair had the shorter path. The instructions included the statement, "You will have only 10 seconds to decide, so please make a snap judgment—there will be no time to make a detailed comparison." On completion of the procedure, participants were asked what characteristics of a tour, if any, they used in making a judgment.

RESULTS AND DISCUSSION

Crossings. The first result of interest was whether a tour with a crossing was more likely to be judged as longer than its matched counterpart with no crossing. Of the 20 comparisons involving crossings and non-crossings, the number of times participants selected the option with a crossing as longer ranged from 3 to 20, with a mean of 13.20 (66%) and standard deviation of 4.96. The difference from the expected mean of 10 was significant by a one-sample t-test, $t(19) = 2.81$, $p = .01$, $d = 0.63$. A Bayesian analysis indicated a Bayes Factor in favor of the alternate hypothesis of $BF_{10} = 4.71$, interpretable as representing "substantial" evidence for the alternate hypothesis (Jarosz & Wiley, 2014). There was little evidence of a difference between the 10-node and 20-node problems in the tendency to judge stimuli with crossings as longer, the means (and standard deviations) being 6.35 (2.69) for the former and 6.85 (2.54) for the latter, $t(19) = 1.31$, $p = .20$. The Bayes Factor of $BF_{10} = 2.04$ indicated "weak" evidence for the hypothesis of a difference (Jarosz & Wiley, 2014). Fifteen of the 20 participants (75%) cited crossings as a factor in their judgment, indicating that crossings were not only noticed, but also recognized as a factor in decisions making. Thirteen of these participants considered crossings to signify longer tours, two as indicating shorter tours. The results supported the hypothesis that tours with crossings are judged to be longer than those without.

Areas. Ten of the comparisons involved pairs with no crossings but different areas, with differences ranging from 3% to 46%. Participants selected the tour with the smaller area as the shorter from three to nine times of 10 decisions, with a mean of 6.20 (62%) and standard deviation of 1.60. The difference from the expected mean of 5 was significant by a one-sample test, $t(19) = 3.27$, $p < .01$, $d = 0.73$, $BF_{10} = 11.00$, indicating "strong" evidence in favor of the alternate hypothesis. There was no difference between the 10-node and 20-node problems in judging smaller areas to be shorter, the means (and standard deviations) being 3.20 (0.93) for the former and 3.00 (1.00) for the latter, $t(19) = 0.81$, $p = .43$. There was marginal evidence in favor of the null hypothesis, with $BF_{01} = 3.22$. There was a positive but non-significant correlation between the number choosing the smaller area as shorter and the percentage difference between the areas of the stimulus pairs, $r = 0.36$, $t(8) = 1.10$, $p = 0.30$, $BF_{10} = 0.62$). Five of the 20 participants mentioned area and one cited convexity as relevant factors in their judgments. Of these six, five stated that smaller areas indicated shorter paths, while one did not specify a direction. The performance results supported the second hypothesis, that paths with smaller areas are judged to be shorter.

Overall, a majority of participants (75%) cited crossings as a factor in their judgments, while a minority (30%) mentioned area or convexity, suggesting that of the two, crossings were more noticeably or actively linked by participants to their decision making. Nevertheless, the approximately equal effect sizes imply that the two factors had a similar impact on length judgments.

The effect of crossings on judgments may arise from the same cognitive sources as a *crossing avoidance* strategy (van Rooij et al., 2003), which proposes that crossing avoidance may be used during tour production as a means of generating short tours. If crossing avoidance leads to better tours then, arguably, the more crossings that are avoided, the better a tour is likely to be. If so, and if judgments are associated

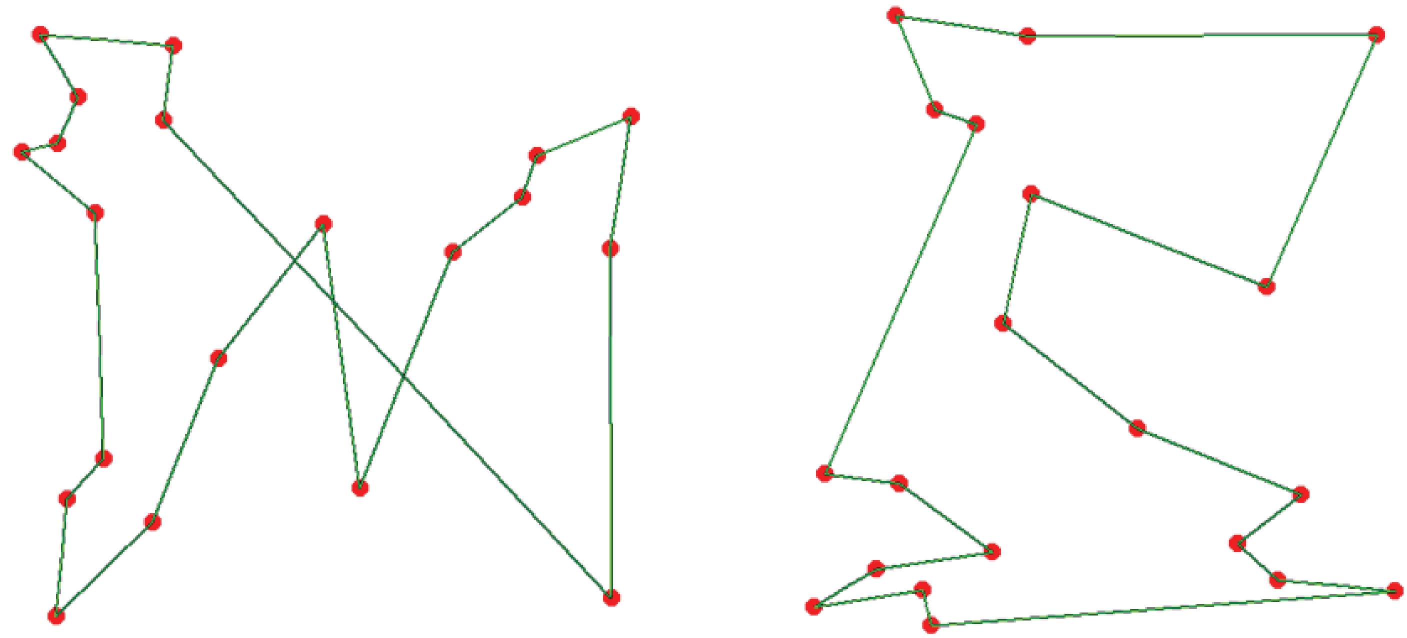

Figure 1.
Example of a stimulus pair used in Experiment 1.

with the same strategy, then tours with fewer crossings should be judged shorter than those with more. Experiment 2a was designed to investigate this.

Although the results showed a significant tendency for the stimulus with the smaller area to be judged the shortest of a pair, some doubts remain about a direct connection between the two factors. For one thing, only a minority of participants identified area or convexity as a factor. In addition, there was no significant correlation between actual area differences and a tendency to choose the smaller area as shorter. One question is whether people can actually reliably detect the differences used here in the areas of these complex polygons. If not, then presumably some related but more detectable stimulus characteristic was used in judging comparative stimulus length. To further explore this issue Experiment 2b investigated whether people can reliably judge differences in area with this type of stimulus.

EXPERIMENT 2A

Experiment 2a was conducted to further examine the influence of crossings on comparative judgments of tour length. In the first experiment, a stimulus tour with a crossing was more likely to be judged longer than a tour of the same length without a crossing. The present experiment examined whether the mere presence of a crossing accounts for the whole effect, or whether more crossings in a single tour add to the effect.

In addition, Experiment 2a introduced a procedural modification. In the first experiment some anomalies arose between participants' decisions and their subsequently expressed strategies. For example, of the 13 stating that crossings were a sign of a longer tour, three selected a tour with a crossing as *shorter* in the majority of comparisons, in one case, 85% of the time. While there are several reasons why this may have occurred, a simple possibility is that the way judgments were recorded could have been a factor. Each member of a comparison pair was labeled A or B, and participants indicated their choice by reading out the label, which was then recorded by the experimenter. It is possible that during the session a participant may have confused whether they were to report the shorter or the longer member of a pair, or some other reporting or recording error may have taken place. Experiment 2a aimed to improve the procedure by using a more objective indicator of choice, and one that was better aligned with the direction of the judgment.

METHOD

Participants. Participants were 20 new volunteers recruited from the campus community at the University of Victoria.

Stimuli. The initial base stimuli were five 20-node point sets each with a completed E-TSP tour with no crossings. Four of the stimuli were taken from the first experiment and one was randomly generated for the present experiment. For each point set two additional tours were randomly generated, one with one crossing and one with more than one crossing, under the constraint that tours with crossings were (1) shorter than the corresponding base tour, and (2) within 2% of its length.

Each of the five stimulus triads yielded three paired comparisons, 0 crossing versus 1 crossing, 0 crossing versus >1 crossing, and 1 crossing versus >1. The resulting 15 paired comparisons were presented as Powerpoint slides in random order. On each slide, a randomly selected member of the pair was rotated through 180 degrees to obscure the fact that members of a pair involved the same point set. Figure 2 provides an example of a zero versus multiple crossing comparison used in the experiment.

Procedure. The procedure was the same as in the first experiment except that, having made a judgment, participants were required to indicate it by clicking on the picture frame of the perceived shorter path and dragging the image frame until the path appeared to be the same length as the comparison tour. This provided an objective record of which tour was judged the smaller and served as a reminder to participants that the decision was to select the shorter of the two tours. In the present experiment, participants were not asked about strategies following the procedure.

RESULTS AND DISCUSSION

The mean number (and standard deviation) of times that stimuli were judged shorter were as follows. For the 0 v 1 crossing comparison, stimuli with no crossings were judged shorter in 3.35 (s = 1.09) of 5 comparisons (67%). For the 0 v > 1 crossing comparisons, the stimuli with no crossings were judged shorter on 3.55 (s = 1.32) occasions (71%). Finally, for the 1 v > 1 comparisons, stimuli with fewer crossings were judged shorter 2.3 times (s = 0.98), or 46%. Results were analyzed using a repeated measures analysis of variance, indicating an overall significant difference among means, $F(2,38) = 7.05$, $p < .01$, MSe = 1.28, $\eta^2_p = .27$, $BF_{10} = 43.66$. Bonferroni post hoc comparisons showed a significant difference for the 0 v 1x comparison, $p < .05$, $d = .82$, $BF_{10} = 5.77$ and the 0 v > 1x comparison, $p < .05$, $d = .98$, $BF_{10} = 9.76$, but no difference when both stimuli had crossing(s), p = 1.00, $BF_{10} = 0.28$.

The results appeared to support and extend those of the first experiment. First, tours with a crossing were systematically judged to be longer than tours with no crossing, replicating the finding of Experiment 1. Second, and consistent with this, tours with more than one crossing were judged to be longer than tours with none. Third, the effect sizes in the 0 v 1x and 0 v > 1x comparisons were relatively similar, at d = .82 and .91, respectively, suggesting that the mere presence of a crossing had as much impact on judgments as multiple crossings. Finally, and supporting the latter conclusion, when participants compared tours having one crossing with tours having more than one, there was no significant difference in which was judged to be longer.

The results support those of the first experiment in concluding that people use a crossings heuristic in deciding which of two tours is longer.

EXPERIMENT 2B

Experiment 2b was conducted to further examine the influence of area on judgments of tour length. In the first experiment, a tour with a smaller area was more likely to be judged as shorter than a tour of greater area, and it was proposed that people may use area differences as a heuristic for judging perimeter lengths. However, only a minority of the participants in the first experiment (30%) cited area or convexity explicitly as a factor in their decisions, suggesting that they may not have been aware of area differences. This raises the issue whether people can in fact reliably discriminate differences in areas of these complex polygons. If they cannot, then it suggests that some other factor was responsible for the results of the first experiment. The present experiment was conducted with two main aims. One was to replicate the area-related findings of Experiment 1. The other was to test whether participants can indeed reliably judge relative differences in areas in these stimuli.

METHOD

Participants. The participants were the 20 volunteers from Experiment 2a, and they took part in the present procedure immediately following the previous one.

Stimuli. The initial base stimuli were six 10-node point sets and six 20-node point sets, each with a completed E-TSP tour with no crossings and with a tour length in the range typical of human solutions (from 0% to 25% above the optimal tour). Four of the stimuli were taken from the first experiment and eight were generated for the present experiment. For each point set a second tour was found of approximately the same length (within 2%) but with a different area (> 5%). Length differences between pair members ranged from 0.04% to 1.24% with a mean of 0.30%. Area differences ranged from 8% to 84% with a mean of 34%.

Two of the stimulus pairs represented special cases. For these, the optimal tour was used as the test stimulus. Then, in each case, an alternative quasi-optimal tour (within 0.04%) was found using the procedure described in MacGregor (2012). One of these pairs had an area difference of 9%, the other, of 22%.

The result of the process was 12 paired comparisons which, as previously, were presented in random order as Powerpoint slides. As before, a randomly selected member of each pair was rotated through 180 degrees to obscure the fact that both tours in each pair involved the same point set. An example is shown in Figure 3.

Procedure. In this case there were two separate components to the procedure that were conducted one after the other, in one case requiring participants to select the shorter member of a pair, in the other, to select the one with the smaller area. One

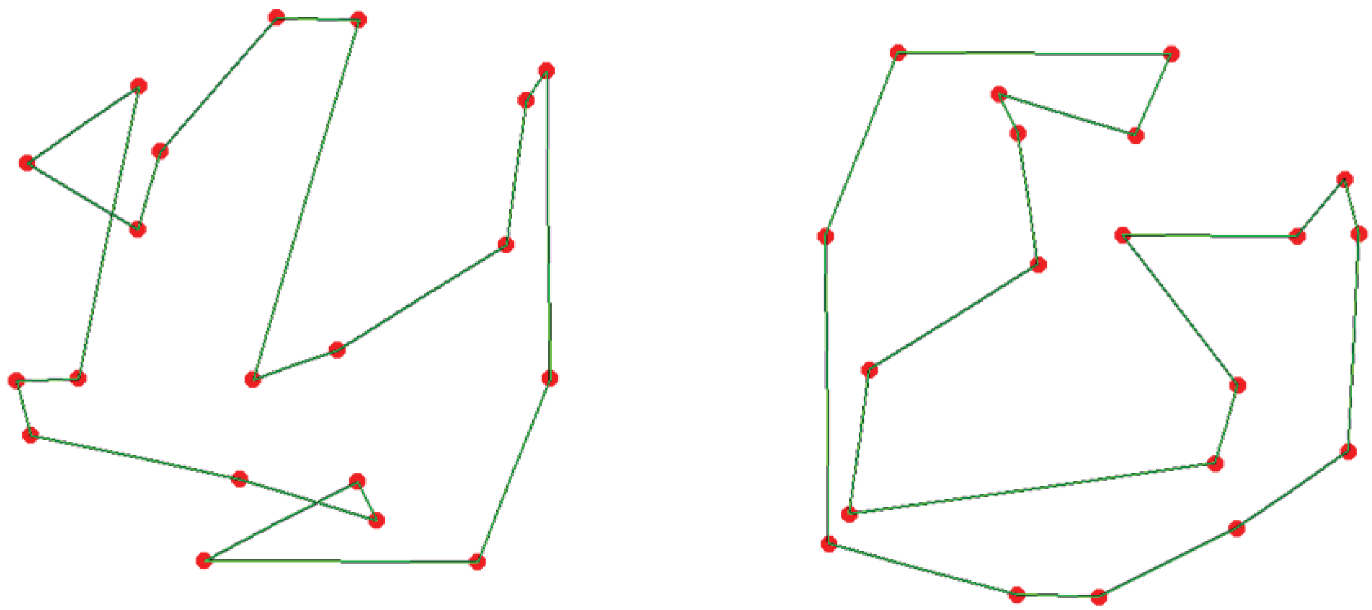

Figure 2.
Sample stimulus from Experiment 2a showing tours with two crossings (left) and no crossings (right).

half of the participants made the length judgments first followed by area judgments of the same stimuli (but in a different random order). The remaining participants made length and area judgments in the reverse order. Otherwise, the procedure was the same as in Experiment 2a.

RESULTS AND DISCUSSION

Length judgments. The number of times that pair members with the smaller area were judged shorter ranged across participants from 3 to 12 of 12 judgments, with a mean of 7.90 (s = 2.29), representing 66% of decisions. The difference from the expected mean of 6 was significant by a one-sample test, $t(19) = 3.71$, $p < .01$, $d = 0.83$, $BF_{10} = 25.81$. There was no significant difference between the 10-node and 20-node problems in judging smaller areas to be shorter, the means (and standard deviations) being 3.95 (1.47) for the former and 3.95 (1.15) for the latter. As in Experiment 1, there was a positive but non-significant correlation between the number choosing the smaller area as shorter and the percentage difference between the areas of the stimulus pairs, $r = 0.15$. The results replicated those of Experiment 1 in showing that participants appear to use differences in area (or a correlated attribute) in judging comparative lengths of tours. At the same time, the lack of any significant correlation between the percentage differences in area and the strength of tendency to choose the smaller area as shorter is somewhat surprising. Other factors being equal, larger area differences should be more apparent than smaller ones and therefore more available for use in length judgments. However, it is possible that even the lower area differences were sufficiently above threshold that additional degrees of difference did not have an effect.

Area judgments. For each of the 12 stimulus pairs, participants judged which member of a pair had the smaller area. The number of correct decisions ranged across participants from 9 (75%) to 12 (100%), with a mean of 10.90 (s = 1.07). Overall, 91% of judgments were correct, indicating that participants were highly accurate in detecting differences in area between pairs of these complex polygons. In this case, there was a significant correlation between the percentage difference in area between members of a pair and the number of correct decisions, $r = 0.64$, $t(10) = 2.81$, $p < .05$, $BF_{10} = 3.27$.

GENERAL DISCUSSION

The article reported several studies designed to test the proposal that in judging the comparative lengths of completed E-TSP tours, people do what they appear to do in attempting to construct tours, which is to apply heuristics. The studies tested two possible heuristics, the presence of crossings, and differences in area (or convexity). The first is related to proposals that, in constructing tours, people avoid creating crossed arcs (van Rooij et al., 2003), the second, to models that propose that they use a form of convex hull (Best, 2005; MacGregor, Ormerod, & Chronicle, 2000).

The results of the first experiment showed that people systematically judge tours with crossings to be longer than tours without. Furthermore, a majority of participants were aware of using this characteristic in making their judgments (75%). The results of the second experiment supported and extended these results by finding again that tours with crossings were judged to be longer but that those with more crossings were not judged to be longer than those with only one.

In addition, the results of the first experiment indicated that tours enclosing a smaller area were judged to be the shorter of two tours of equivalent lengths, although participants were less likely to be aware of this characteristic influencing their judgments (30%). Participants mentioned a variety of other stimulus features that may have been related to area or convexity but which were more difficult to interpret unambiguously. These included complexity, convolutions, number of angles, number of zigzag lines, and number of long, straight, lines. It is possible that some, or all, of these represented an attempt to articulate stimulus features that are correlated with area. The second experiment replicated the finding that tours with smaller areas were judged shorter. It also established that participants were able to reliably detect differences in areas of these magnitudes, thereby confirming that area differences could provide a feasible heuristic for making length comparisons.

Carruthers (2015) proposed that when problem solvers are faced with a problem where it is infeasible to identify whether or not the goal has been reached, they may modify the goal into one that is feasible to identify. In the case of the E-TSP, an example of a possible modified goal is "a valid tour" (p. 49). The fact that people appear to use heuristics both in constructing tours and, as demonstrated here, in comparing their lengths, does not necessarily challenge this view, as heuristically guided decisions could be regarded as a form of goal modification. That is, if "a valid tour" represents a modified goal, then "a valid tour with no crossings" and "a valid tour with a small area" might be regarded as more specifically modified goals. If this interpretation is valid, then identifying the heuristics applied is equivalent to "identifying the problem as encoded." This would also imply that tour construction is driven by such strategies, and whether tours with no crossings or small areas are the result of these strategies or are the outputs of other heuristics remains an open question. Nevertheless, the present findings related to crossings could be interpreted as consistent with a crossing-avoidance hypothesis (van Rooij et al., 2003), while similarly the findings related to area are consistent with models based on the convex hull (Best, 2005; MacGregor et al., 2000).

In the present experiments, although the paired comparisons involved tours of virtually identical lengths, one tour was always slightly shorter than the other. Thus, there was always a "correct" answer. However, because the difference was apparently below threshold, participants' decisions were as often incorrect as correct. Heuristic decision making has sometimes been characterized as "irrational" if it leads to incorrect outcomes (Kahneman & Tversky, 1972), and the present decisions appear to meet this description. However, this could be the result of the specific stimuli used and it may be that the present heuristics have some ecological validity.

As far as I am aware, the present study is the first to systematically examine stimulus factors used in judging the lengths of E-TSP tours and, more generally, the perimeter lengths of complex polygons. As such, it is exploratory rather

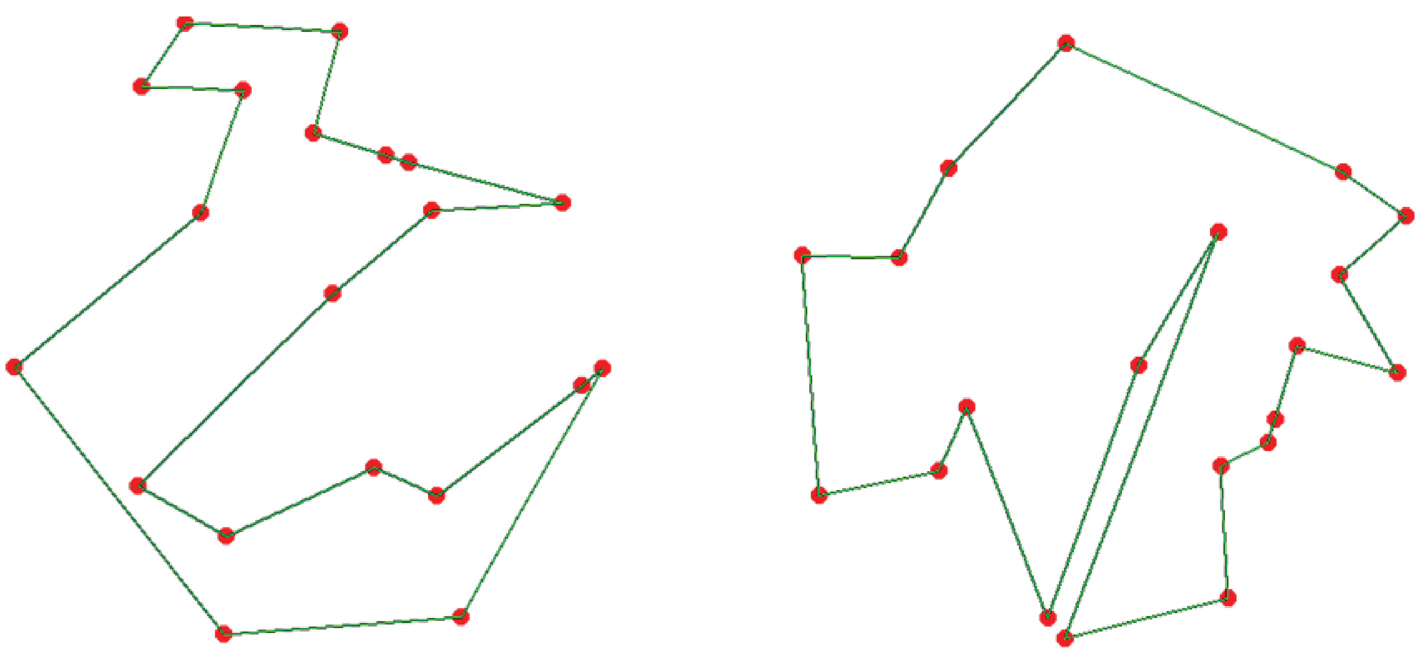

Figure 3.
A sample stimulus from Experiment 2b showing tours of equal length but different areas.

than exhaustive, and provided a very limited manipulation of factors. One limitation is that it did not take into account that there are different types of possible intersections (van Rooij et al., 2003) and the question remains open as to whether some types of crossings are more likely to affect judgments than others. Another limitation is that in controlling both point sets and perimeter lengths the present comparisons confounded area, convexity, and circularity. Having now established that at least one of these factors is relevant, future experiments may be designed to identify which are critical. Finally, there are stimulus factors other than those studied here that may influence length judgments.

REFERENCES

Best, B. J. (2005). A model of fast human performance on a computationally hard problem. *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 256–261). Mahwah, NJ: Lawrence Erlbaum.

Dry, M. J., Lee, M. D., Vickers, D., & Hughes, P. (2006). Human performance on visually presented traveling salesperson problems with varying numbers of nodes. *Journal of Problem Solving*, *1*, 20–32.

Dry, M. J., & Fontaine, E. L. (2014). Fast and efficient discrimination of traveling salesperson problem stimulus difficulty. *Journal of Problem Solving*, *7*(1). 84–93. https://doi.org/10.7771/1932-6246.1160

Carruthers, S. L. (2015). *The role of the goal in problem solving hard computational problems: Do people really optimize?* (Unpublished doctoral dissertation). University of Victoria, Victoria, British Columbia.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Graham, S. M., Joshi, A., & Pizlo, Z. (2000). The traveling salesman problem: A hierarchical model. *Memory & Cognition*, *28*, 1191–1204.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *Journal of Problem Solving*, *7*, 2–9. https://doi.org/10.7771/1932-6246.1167

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

MacGregor, J. N. (2012). Indentations and starting points in traveling sales tour problems: Implications for theory. *Journal of Problem Solving*, *5*(1), 2–17. https://doi.org/10.7771/1932-6246.1140

MacGregor, J. N., & Ormerod, T. (1996). Human performance on the traveling salesman problem. *Perception & Psychophysics*, *58*, 527–539.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (1999). Spatial and contextual factors in human performance on the traveling salesman problem. *Perception*, *28*, 1417–1427.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2000). A model of human performance on the traveling salesperson problem. *Memory & Cognition*, *28*(7), 1183–1190.

Ormerod, T. C., & Chronicle, E. P. (1999). Global perceptual processing in problem-solving: The case of the traveling salesperson. *Perception and Psychophysics*, *61*, 1227–1238.

Philip, J. (2004). The area of a random convex polygon. TRITA-MAT-04-MA-07, June. Retrieved from http://www.math.kth.se/~johanph/area12.pdf

Simon, H. A. (1983). Discovery, invention, and development: Human creative thinking. *Proceedings of the National Academy of Science, USA*, *80*, 4569–4571.

Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

van Rooij, I., Stege, U., & Schactman, A. (2003). Convex hull and tour crossings in the euclidean traveling salesperson problem: Implications for human performance studies. *Memory & Cognition*, *31*, 215–220.

Vickers, D., Lee, M. D., Dry, M. J., Hughes, P., & McMahon, J. A. (2006). The aesthetic appeal of minimal structures: Judging the attractiveness of solutions to traveling salesperson problems. *Perception & Psychophysics*, *68*(1), 32–42. https://doi.org/10.3758/BF03193653