

## Individual Differences in Optimization Problem Solving: Reconciling Conflicting Results

*Edward P. Chronicle<sup>1</sup>, James N. MacGregor<sup>2</sup>, Michael Lee<sup>3</sup>,  
Thomas C. Ormerod<sup>4</sup> and Peter Hughes<sup>5</sup>*

### Abstract

Results on human performance on the Traveling Salesman Problem (TSP) from different laboratories show high consistency. However, one exception is in the area of individual differences. While one research group has consistently failed to find systematic individual differences across instances of TSPs (Chronicle, MacGregor and Ormerod), another group (Vickers, Lee and associates) has found individual differences both within TSP performance and between TSP performance and other cognitive tasks. Among possible reasons for the conflicting results are differences in procedure and differences in the problem instances used. To try to resolve the discrepancy, we collected data on TSP performance by combining the procedure used by one group with problem instances used by the other. The comparison involved nine 30-node and nine 40-node TSP problems previously used by the Vickers group, using computer presentation. Here, we had the same problems completed by 112 participants using a paper-and-pencil mode of presentation. We examined the results in the form of distributions of correlations across individuals for each pair of problems of the same size. The distributions for the computer and paper forms of presentation were very similar, and centered between correlations of 0.20 and 0.30. The results indicated the presence of individual differences at a level that fell between those previously reported by the two laboratories. The pattern of results indicated that previous discrepancies did not arise because of differences in procedure. Instead, individual differences appeared to become more prevalent as the difficulty of problems increased. The results are consistent with an explanation that performance on simpler instances is

<sup>1</sup> Department of Psychology, University of Hawaii, USA.

*Ed Chronicle was the initiator of the research reported here but, sadly, did not live to see its completion. The Journal of Problem Solving published a tribute to Ed in 2007. —James MacGregor*

<sup>2</sup> School of Public Administration, University of Victoria, BC, Canada; <sup>3</sup> Department of Psychology, University of California, Irvine, California, USA; <sup>4</sup> Department of Psychology, Lancaster University, UK; <sup>5</sup> Department of Psychology, University of Adelaide, Australia

Address correspondence to:

James N. MacGregor, School of Public Administration, University of Victoria, PO Box 1700 STN CSC, Victoria, BC V8W 2Y2, Canada

Email: [jmacgreg@uvic.ca](mailto:jmacgreg@uvic.ca); Telephone: 250-721-6435

dominated by lower-level processes, but that as instance difficulty increases, higher-level functions become increasingly involved.

## Introduction

In the area of problem solving there is growing interest in how people perform on instances of combinatorial optimization tasks. An example is the classic Traveling Salesperson Problem (TSP), where the goal is to find the shortest route through a set of points and return to the origin. The TSP is a member of a class of mathematical problems that have so far proven to be computationally intractable. That is, no general algorithmic procedure has been found that is guaranteed to solve in polynomial time. Instead, heuristic procedures have been developed that can produce very good solutions in reasonable computational time, but that cannot guarantee finding the optimum (Applegate, Bixby, Chvátal, & Cook, 2006).

Interestingly, human solutions are often better than those of the simpler heuristic procedures (Graham, Joshi, & Pizlo, 2000; MacGregor & Ormerod, 1996). In addition, while the computational times per node for successful heuristic procedures typically increase as a function of the number of nodes, it seems that human solution times per node remain constant. In other words, human solution times per problem increase in proportion to the number of nodes (Graham et al., 2000; Dry, Lee, Vickers, & Hughes, 2006; Pizlo, Stefanov, Saalweachter, Li, Haxhimusa, & Kropatsch, 2006). Other generally agreed-upon findings are that human solutions are typically close to optimal (Graham et al., 2000; MacGregor & Ormerod, 1996; van Rooij, Schactman, Kadlec, & Stege, 2006; Vickers, Butavicius, Lee, & Medvedev, 2001) and rarely self-intersect (MacGregor, Ormerod, & Chronicle, 2000; van Rooij, Stege, & Schactman, 2003; Vickers, Lee, Dry, & Hughes, 2003).

While the results from different laboratories show substantial agreement on several basic aspects of performance, there is one fundamental issue on which they diverge. This is the question of whether or not performance differs reliably across individuals. While Chronicle, MacGregor and Ormerod have consistently failed to find individual differences in TSP performance (Chronicle, MacGregor & Ormerod, 2006; MacGregor & Ormerod, 1996), Lee, Vickers and colleagues have reported significant differences, not only within TSP performance (Vickers, Bovet, Lee, & Hughes, 2003; Vickers et al., 2001), but between TSP performance and other cognitive tasks (Burns, Lee, & Vickers, 2006; Vickers, Mayo, Heitman, Lee, & Hughes, 2004).

Table 1 provides a representative sample of the relevant findings. The table shows the source of data, the number of participants, the number of TSP problems employed ( $k$ ), the number of nodes, and the average correlation between path lengths across all  $k(k-1)/2$  pairs of problems. As can be seen from the final column, there is a striking difference in the results from the two laboratories.

Chronicle, MacGregor and Ormerod (2006) proposed that procedural differences might, in part, explain these apparent inconsistencies. They argued that features of the

Vickers, Bovet, et al. (2003) procedure, for example, may have allowed learning by providing feedback after each trial, and that the results may have reflected individual differences

**Table 1.** Individual differences in TSP performance reported by two different laboratories.

Source	N	Number of TSPs	Number of nodes	Mean correlation
MacGregor&Ormerod (1996), Exp. 1	45	6	10	$\bar{\rho} = .06$
MacGregor&Ormerod (1996), Exp. 2	20	7	20	$\bar{\rho} = .02$
Chronicle et al (2006), Exp. 2	24	6	15	$\bar{\rho} = .10$
Chronicle et al (2006), Exp. 3	18	8	15	$\bar{\rho} = .01$
Vickers et al (2001), Exp. 1, Group O	18	6	10,25,40	$\bar{\rho} = .87$
Vickers, Bovet et al (2003)	81	2 <sup>1</sup>	30	$r = .61$
Vickers et al (2004), Exp. 1	69	5	50	$\bar{r} = .53$

<sup>1</sup>Two different sets of 6 TSPs were used. Scores were averaged within each set and the correlation computed between the two sets of average scores.

in learning ability. However, while this could explain the results reported in Vickers, Bovet, et al. and Vickers, Lee, et al. (2003), it cannot explain those found in studies where feedback was not provided (Vickers et al., 2001; Vickers, Lee, Hughes, Dry & McMahon, 2006). In contrast, Vickers, Lee and colleagues have argued that the failure to find individual differences by MacGregor and Ormerod (1996) and Ormerod and Chronicle (1999) is likely to have arisen from ceiling effects, due to the simple and highly constrained nature of the stimuli used (Vickers et al., 2001; Vickers et al., 2006). Again, however, while this conjecture may explain some findings, it cannot explain them all. Chronicle et al. (2006), in Experiment 2, using unconstrained random stimuli, failed to find individual differences. Also, while the problems were relatively small (15 nodes), there was no evidence of ceiling effects. In fact, mean path lengths were very similar to those reported in Vickers, Bovet et al., who used similar problems and found significant individual differences.

To try to resolve these puzzling discrepancies, the two laboratories decided to pool resources by conducting an empirical study using the procedures of Chronicle and colleagues combined with stimuli previously used by Vickers and colleagues and compare the results with those found by the Vickers group. The present article is a result of this collaboration.

## Method

The comparison used nine 30-node and nine 40-node TSP problems. The data for the computer presentation mode came from previous experiments conducted by the Vickers group. Specifically, data for the 30-node problems came from Vickers, Bovet, et al. (2003), and for the 40-node problems, from Vickers, Lee, et al. (2003). Details of procedures may be found in the original publications. In both cases, participants were tested individually,

problem stimuli were presented by computer and node-to-node connections were indicated by participants through pointing-and-clicking. After completion of each problem, feedback was provided by visually displaying each participant's solution with the optimal solution superimposed, and by displaying the participant's versus the optimal path length. Eighty-one participants completed the 30-node problems. Ninety-three participants completed the 40-node problems.

The same problems were completed by 112 students of the University of Hawaii, working in a group setting. Problem stimuli were presented on paper and participants drew their solutions. No feedback was provided.

## Results and Discussion

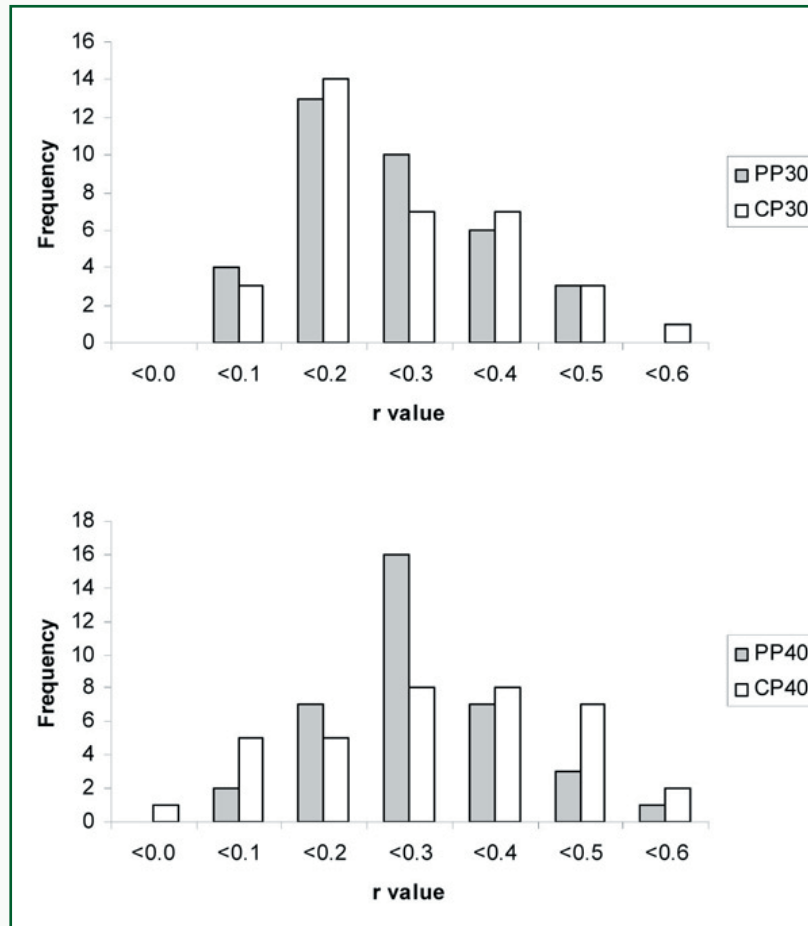
Because of the many uncontrolled differences between the computer presentation (CP) and paper presentation (PP) groups, we have adopted an exploratory, descriptive approach to the results. The main question of interest is whether there was greater evidence for reliable individual differences in performance in the computer presentation group than the paper presentation group. For both, performance was measured as the difference between the length of path produced by a participant and the optimal path length, expressed as a percentage of the optimal length (percentage above optimal, or PAO). We calculated the Pearson correlation coefficients in PAO between all pairs of 30-node problems for the computer group and for the paper group. That is, we calculated the  $9 \times 8 / 2 = 36$  correlation coefficients corresponding to every possible combination of the 30-node problems for the computer group, and then did the same for the paper group. We repeated this for the 40-node problems, again resulting in 36 correlation coefficients. The distributions of these sets of correlation coefficients are shown in Figure 1 below. The upper and lower panels show the results for the 30-node and 40-node problems, respectively.

Perhaps the most striking feature of the results displayed in Figure 1 is the similarity between the PP and CP groups. If the results were consistent with those shown in Table 1, we might expect the PP distributions to be centered around an  $r$  value of less than 0.10, and the CP distributions to be centered around a value greater than 0.50, with little or no overlap between the two distributions. In contrast, all four of the distributions shown in Figure 1 have a central tendency between 0.20 and 0.30, and clearly overlap to a considerable degree.

The results provide little or no evidence that mode of presentation affected individual differences in performance. Instead, they seem to indicate higher levels of individual differences than previously found by the Chronicle group and lower levels than reported by the Vickers laboratory. The latter result is probably due to the fact that the present correlations were based on raw performance scores and not on the average of individual scores across problems of the same size. The raw-score method was typically used by the Chronicle

group. In contrast, the Vickers group typically obtained average scores for individuals and used these as the basis for calculating correlations, which would normally result in

**Figure 1.** Frequency distributions of pair-wise correlation coefficients on 30-node (upper panel) and 40-node (lower panel) TSPs, for paper presentation (PP) versus computer presentation (CP).



higher correlation coefficients. The question remains as to why the correlations observed here are higher than those previously found by the Chronicle group. A possible explanation is proposed below.

#### ***Effects of problem difficulty***

For the paper presentation mode, the present problems were larger than previously used in testing for individual differences. This raised the possibility that individual differences in performance became more reliable as problem difficulty increased.

One approximate indicator of problem difficulty is the number of nodes: Other things being equal, the 40-node problems should be somewhat more difficult than the 30. Were individual differences in the paper group greater in the former case? The results suggest

that they may have been. The pair-wise correlations between 40-node problems ranged from 0.08 to 0.52 with a mean of 0.26. For 30-node problems, they ranged from 0.04 to 0.45, with a mean of 0.22. (Fisher's z-transform was used in averaging.)

A more sensitive measure of problem difficulty is provided by the actual performance scores. However, since these were used in the calculation of individual differences, a problem of lack of independence arises. To avoid this, we used the PAO scores from the computer group as the indicator of problem difficulty.

First we divided the 18 problems into 3 groups of 6, in order of increasing PAO. (The groups had mean PAO of 5%, 7% and 8%, respectively). Next we used the results from the paper condition to calculate the pair-wise correlations between problems in each of the three groups. For the easiest group, the resulting correlations ranged from 0.04 to 0.40, with a mean of 0.20, for the next, from 0.06 to 0.40, with a mean of 0.25, and for the most difficult group, from 0.12 to 0.36, with a mean of 0.26. Again, the results were consistent with individual differences increasing with problem difficulty.

Finally, we computed an individual differences index for each problem as the average pair-wise correlation between a problem and each of the other 17 problems. These ranged from 0.15 to 0.29. The correlation between this index and a problem's mean PAO was 0.42, across the 18 problems. Again, the result was consistent with the hypothesis that individual differences become increasingly reliable as problem difficulty increases.

## General Discussion

The purpose of the research was to reconcile, if possible, conflicting results on individual differences in performance on TSPs. One laboratory, using relatively small problems and a paper-and-pencil approach, has consistently reported no, or small, individual differences in performance, while another, using a wider range of problem sizes combined with computer presentation, has consistently reported larger, reliable, individual differences. The present research combined moderately large problems generated and tested by the latter laboratory with the data-collection procedures of the former laboratory. The results indicated the presence of individual differences at a level that falls between those previously reported by the two laboratories. The pattern of results indicated that individual differences become more prevalent as problem difficulty increases.

There are at least two possible explanations for this apparent relationship between problem difficulty and individual differences. One is that it is an artifact, caused by ceiling effects. That is, at lower levels of difficulty, range restrictions artificially constrain individual differences. If this is the case, then we would expect less variability in performance among the simpler group of problems than among the more difficult. The standard deviations in PAO for the three groups of problems examined here were 0.9%, 0.6% and 0.6%, from simplest to most difficult, respectively, showing no indication of greater range restriction

in the simpler group.

A second explanation is that the result reflects differences in solution processes as problem difficulty increases. A number of theoretical accounts have suggested that how people solve TSPs is, in part, based on perceptual processes (Graham et al., 2000; MacGregor et al., 2000). It seems plausible that, for very simple problems, perceptual processing predominates but, as problem difficulty increases, more analytical processes come into play. An analogous phenomenon might be the switch from subitizing to counting in enumerating object displays. With few objects, subitizing provides a quick and accurate enumeration but, as the number of objects increases, accurate enumeration requires algorithmic counting (Kaufman; Lord; Reese & Volkman, 1949).

The conclusion that performance on TSPs is determined both by perceptual and cognitive/analytical processes is consistent with the findings of Burns et al. (2006) and van Rooij et al. (2006). Burns et al. tested 101 subjects on a battery of combinatorial optimization problems, including 30-, 60- and 90-node TSP problems, and on standard psychometric tests designed to measure various cognitive abilities. These authors found significant correlations between TSP performance and measures of both low-level perceptual abilities, particularly involving spatial relations, and more high-level cognitive abilities, particularly involving verbal analytic reasoning. While all correlations were significant, those of verbal analytic reasoning with TSP performance were higher than those of visuospatial ability. Given the large sizes of the TSPs used, this appears to be consistent with the present explanation. Burns et al. advocated a structural equation model as a parsimonious explanation of the intercorrelations between tests they observed. This structural equation model explained individual variation in TSP performance both in terms of a fluid intelligence factor and a visuospatial intelligence factor. Fluid intelligence can be thought of as a measure of abstract reasoning ability, and visuospatial intelligence can be regarded as a measure of the ability to manipulate mental representations of shapes, forms and positions of objects. The possibility raised by the current results, that solving TSP problems with lower levels of complexity will tend to rely on visuospatial ability, while solving problems of greater complexity will tend to rely on abstract reasoning ability, is entirely consistent with the mixture observed across a wide range of problem sizes considered in the Burns et al. study.

This conjecture is also consistent with the findings of van Rooij et al. (2006), who compared the TSP performances of three age groups (adults, 12-year-olds and 7-year-olds) on 15 randomly generated problems, 5 each of  $n = 5, 10$  and  $15$ . They concluded that their results with the younger children indicated that high-level TSP performance is attainable with perceptual processing alone, but also noted that adult participants used cognitive-analytic skills to improve upon what their perceptual systems provided.

The hypothesis that individual differences increase as TSP problems become more complex, because of a required shift from basic perceptual mechanisms to higher-order

cognitive or analytic mechanisms, is one that is open to empirical investigation. It raises the basic question of how the “complexity” of a TSP problem should be assessed. In our analysis, we have used the number of nodes as a crude index of complexity, but there are surely more sophisticated approaches that should be developed and explored. If complexity does depend in part on the structure of people’s perceptual and cognitive representations, then theories of perceptual organization and memory ought to play a leading role in understanding how people solve TSP problems, and when and why their solutions differ.

## References

- Applegate, D. L., Bixby, R. E., Chvátal, V., & Cook, W. J. (2006). *The Traveling salesman problem: A computational study*. Princeton, NJ: Princeton University Press.
- Burns, N. R., Lee, M. D., & Vickers, D. (2006). Are individual differences in performance on perceptual and cognitive optimization problems determined by general intelligence? *Journal of Problem Solving, 1*(1), 5-19.
- Chronicle, E. P., MacGregor, J. N., & Ormerod, T. C. (2006). Optimizing and “pessimizing”: Human performance with instructional variants of the traveling salesperson problem. *Journal of Problem Solving, 1*(1), 74-82.
- 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*(1), 20-32.
- Graham, S. M., Joshi, A., & Pizlo, Z. (2000). The travelling salesman problem: A hierarchical model. *Memory & Cognition, 28*, 1191-1204.
- Kaufman, E. L., Lord, M. W., Reese, T., and Volkman, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*, 498-525.
- 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. (2000). A model of human performance on the traveling salesperson problem. *Memory & Cognition, 28*, 1183-1190.
- Ormerod, T. C., & Chronicle, E. P. (1999). Global perceptual processing in problem-solving: the case of the travelling salesperson. *Perception and Psychophysics, 61*, 1227-1238.
- Pizlo Z., Stefanov E., Saalweachter J., Li Z., Haxhimusa Y., & Kropatsch W. G. (2006). Traveling salesman problem: a Foveating pyramid model. *Journal of Problem Solving, 1*(1), 83-101.
- Rooij, I. van, Schactman, A., Kadlec, H., & Stege, U. (2006). Perceptual or analytical processing? Evidence from children’s and adult’s performance on the Euclidean Traveling Salesperson problem. *Journal of Problem Solving, 1*(1), 44-73.
- Rooij, I. van, 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., Bovet, P., Lee, M. D., & Hughes, P. (2003). The perception of minimal structures: Performance on open and closed versions of visually presented Euclidean Traveling Salesperson problems. *Perception*, *32*(7), 871-886.
- Vickers, D., Butavicius, M. A., Lee, M. D., & Medvedev, A. (2001). Human performance on visually presented traveling salesman problems. *Psychological Research*, *65*, 34-45.
- Vickers, D., Lee, M. D., Dry, M., & Hughes, P. (2003). The roles of the convex hull and number of intersections upon performance on visually presented traveling salesperson problems. *Memory & Cognition*, *31*(7), 1094-1104.
- Vickers, D., Mayo, T., Heitman, M., Lee, M. D., & Hughes, P. (2004). Intelligence and individual differences in performance on three types of visually presented optimisation problems. *Personality and Individual Differences*, *36*, 1059-1071.
- Vickers, D., Lee, M. D., Dry, M., 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.

Paper submitted on Jan 15, 2008.

Final version accepted on Mar 27, 2008.