

# The aesthetic appeal of minimal structures: Judging the attractiveness of solutions to traveling salesperson problems

DOUGLAS VICKERS, MICHAEL D. LEE, MATTHEW DRY,  
PETER HUGHES, and JENNIFER A. McMAHON  
*University of Adelaide, Adelaide, South Australia, Australia*

Ormerod and Chronicle (1999) reported that optimal solutions to traveling salesperson problems were judged to be aesthetically more pleasing than poorer solutions and that solutions with more convex hull nodes were rated as better figures. To test these conclusions, solution regularity and the number of potential intersections were held constant, whereas solution optimality, the number of internal nodes, and the number of nearest neighbors in each solution were varied factorially. The results did not support the view that the convex hull is an important determinant of figural attractiveness. Also, in contrast to the findings of Ormerod and Chronicle, there were consistent individual differences. Participants appeared to be divided as to whether the most attractive figure enclosed a given area within a perimeter of minimum or maximum length. It is concluded that future research in this area cannot afford to focus exclusively on group performance measures.

There has been recent interest in a class of problems, dealing with combinatorial optimization, that have been hitherto neglected in both experimental and differential psychology (see, e.g., Graham, Joshi, & Pizlo, 2000; MacGregor & Ormerod, 1996; MacGregor, Ormerod, & Chronicle, 2000; Vickers, Butavicius, Lee, & Medvedev, 2001; Vickers, Mayo, Heitmann, Lee, & Hughes, 2004). In particular, attention has focused on visually presented versions of the well-known traveling salesperson problem (Lawler, Lenstra, Rinnooy Kan, & Shmoys, 1985), traditionally referred to as the *traveling salesman problem* (TSP). In the planar Euclidean form of the problem, there are  $n$  interconnected cities, represented by  $n$  nodes, and the task is to construct a pathway that departs from one node to visit each node exactly once, returns to the starting node, and is as short as possible. The number of possible pathways is equal to  $\frac{1}{2}(n - 1)!$ . As a result, when  $n$  is small (e.g., 5), there are relatively few pathways to choose from (in this case, 12). However, when  $n$  becomes even moderately large (e.g., 25), the number of possible pathways becomes so great that a computer evaluating a million possibilities per second would take almost 10 billion years to evaluate them all (Stein, 1989).

Such problems are of interest for a number of reasons. In the problem-solving domain, they occupy a strategic intermediate position between early research in the Gestalt

tradition, in which the solution process could be described as an insightful perceptual reorganization (Duncan, 1959), and recent research, in which the emphasis has been on studying the nature, acquisition, and application of expertise in knowledge-rich real-world domains (e.g., Chi, Glaser, & Farr, 1988). Such problems can often be stated simply and understood readily and lend themselves to systematic laboratory investigation. At the same time, they are representative of many realistic, practical problems, as evidenced by the numerous applications of mathematics and computer science dealing with combinatorial optimization.

In the domain of visual perception, visually presented instances of such problems are of interest because there is some agreement that the impressive human ability to arrive at near-optimal solutions to these problems depends heavily on the use of spontaneously occurring perceptual principles. If so, an examination of performance in such tasks may throw light on the operation of these principles, as well as on their possible interaction with higher level cognitive processes. For example, MacGregor and Ormerod (1996) propose that the TSP task "may happen to parallel what is natural for the perceptual system to do . . . when presented with an array of dots" (p. 537). More specifically, Ormerod and Chronicle (1999) and MacGregor et al. (2000) hypothesized that participants begin by processing the TSP array globally by identifying the best figure fitting the dot array and then use this figure either to construct an attempted solution or to evaluate a solution that is presented to them. According to these authors, the best-fitting global figure is provided by the convex hull of the array. The convex hull of a set of points,  $S$ , is the polygon with the smallest area, enclosing  $S$ , that cannot be crossed by a straight line joining any two points in  $S$

---

This research was supported by Australian Research Council Discovery Grant DP 0211150 to D.V. and M.D.L. Correspondence concerning this article should be addressed to M. D. Lee, Department of Psychology, University of Adelaide, Adelaide, SA 5005, Australia (e-mail: michael.lee@adelaide.edu.au).

(O'Rourke, 1993, pp. 71–73). MacGregor and Ormerod (1996) further motivate this hypothesis by suggesting that the convex hull may provide a two-dimensional illustration of the Gestalt law of *Prägnanz* in the same way that a soap bubble surface has been used to illustrate it in three dimensions (Attneave, 1982).

According to Ormerod and Chronicle (1999), the convex hull hypothesis makes two obvious and testable predictions. The first, which follows from the hypothesized correspondence between solution processes and spontaneous perceptual principles, is that TSP solutions that are closer to optimality should be rated more highly as having *good figure*. The second prediction, which follows from the hypothesis that the salient principle in arriving at solutions is the global perception of a convex hull structure, is that solutions with relatively more points on the convex hull (and hence, fewer *internal* points) will constitute simpler TSPs and be rated as *better* figures.

Ormerod and Chronicle (1999) tested these predictions in an experiment in which participants rated the perceived goodness of five possible solutions to each of five 10-point TSP problem instances. The solutions were constructed in the ranges 0%, 11%–18%, 21%–28%, 32%–37%, and 41%–47% longer than the optimal solution. The five problem instances (plus a standard, 10-point *Dantzig* configuration) also contained between one and five internal nodes that were located inside the convex hull bounding each configuration. Figure 1 shows stimulus arrays representative of those presented by Ormerod and Chronicle (1999, Figure 1), with between one and five internal nodes.

Ormerod and Chronicle (1999) found that the participants judged optimal solutions as being good figures and progressively less optimal solutions as being progressively less good figures. They also found that the participants rated the solutions to problem instances with fewer internal nodes (and hence, relatively more nodes on the convex hull) as better figures. They interpreted these results as evidence for the hypothesis that the finding of near-optimal solutions to visually presented TSP instances is mediated by spontaneous global perceptual organizing processes, similar to that underlying the perception of good gestalts, and based on the ready identification of the convex hull of each array.

Some support for the view that the TSP solution process is strongly influenced by the operation of spontaneous perceptual principles is provided by a study by Vickers et al. (2001). In their first experiment, these authors pre-

sented two groups of participants with six different arrays of irregularly positioned dots, each representing the nodes of a TSP instance: two 10-node, two 25-node, and two 40-node arrays. The participants in the *gestalt* group were instructed to draw a pathway that passed through each of the points just once, that returned to the starting point, and that looked “most natural, attractive, or aesthetically pleasing.” The participants in the *optimization* group were instructed to draw a pathway that passed through each point just once, that returned to the starting point, and that was as short as possible. Although Vickers et al. (2001) found that there was a relatively small (but significant) difference in the average length of the pathways produced by the two groups, there was also a considerable overlap between the pathways produced by the two groups, with several participants from one group producing exactly the same pathways as those from the other.

Despite this agreement, however, there are some difficulties with Ormerod and Chronicle's (1999) conclusion regarding the identification of the shortest pathway with the figure judged to have the highest degree of goodness. There are difficulties also with their conclusion that the apparent relation between perceived goodness and number of interior points provides initial support for the view that the convex hull is an important determinant of perceived goodness. These difficulties all stem from the method used by Ormerod and Chronicle to generate their configurations, which consisted of small angular perturbations and incremental transformations of one initial, regular polygon, in which all points fell on the circumference of a circle.

As the authors themselves make clear, “it is still conceivable that TSP solutions are provided through some other set of cognitive operations, unconnected with a judgment of good figure, and that the relationship with good figure . . . is merely coincidental” (Ormerod and Chronicle, 1999, p. 1230). Indeed, a number of other possibilities, besides solution optimality and number of internal nodes, readily suggest themselves as potential factors contributing to the judgments of figural goodness observed by Ormerod and Chronicle.

First, as is clear from Figure 1, the convex hulls for these five figures describe polygons with close to perfect regularity. Thus, degree of perceived rotational symmetry is a possible contributing factor. Indeed, the circular prototype, with which MacGregor and Ormerod (1996) and Ormerod and Chronicle (1999) started, encloses a given area with a perimeter of minimum length. Circular figures

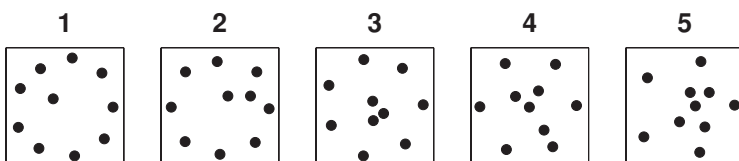


Figure 1. Stimulus arrays representative of those used by Ormerod and Chronicle (1999), with between one and five internal nodes. Based on Ormerod and Chronicle (1999, Figure 1), copyright 1999 by the Psychonomic Society.

have long been regarded as “ideal” figures, and it is possible that the participants in the study by Ormerod and Chronicle judged the presented solutions as aesthetically pleasing to the extent that they approximated the circular prototype used by these authors as a basis for creating successive variants. A measure of circularity is given by  $4\pi A/p^2$ , where  $A$  is the area and  $p$  is the perimeter of a region (Selkirk, 1982, p. 55). Circularity, in this instance, correlates highly with the number of points on the convex hull ( $r = .96$ ), so it is possible that circularity would also provide an equally or more predictive measure of good figure for configurations in which this correlation is less pronounced.

Second, the degree of convexity of the solutions shown in Figure 1 or in Ormerod and Chronicle (1999), as conventionally measured by the ratio between the area enclosed by the solution configuration and the area enclosed by the convex hull (Zunic & Rosin, 2004), varies systematically. Although convexity would be expected to show a correlation with the number of nodes on the convex hull, convexity is potentially a more sensitive measure. (For example, a six-pointed regular star figure can have six convex hull nodes only, but, depending on the depth of the indentations, can possess varying degrees of convexity.) Polivanova’s (1974) participants identified convexity as one of the criteria they employed in arriving at solutions, and convexity is a factor that has long been identified with the differentiation between figure and ground (Kanizsa & Gerbino, 1976; Metzger, 1953), as well as, more recently, with perceptual completion (Liu, Jacobs, & Basri, 1999) and the perception of holes (Bertamini & Croucher, 2003).<sup>1</sup>

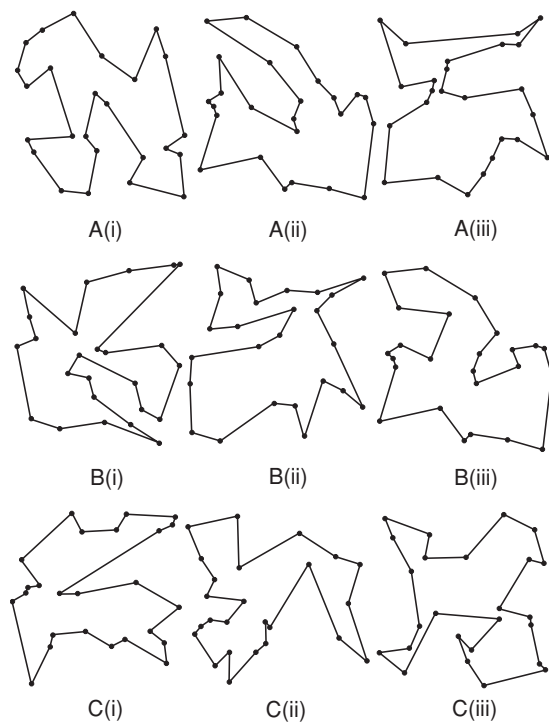
Third, the actual links making up the optimal solutions also vary in regularity, as measured by progressive increases in the standard deviation in length of the connecting edges. (From scanned versions of their figures, we estimate the standard deviations to take values proportional to 0.029, 0.039, 0.111, 0.112, and 0.141 units for the figures with one to five internal nodes, respectively.)

Fourth, the arrays vary with respect to the total number of potential intersections associated with all possible pairs of links between all 10 nodes of each instance. We estimate that the total number of potential intersections, expressed as a percentage of all permissible links, took values of 29.5%, 25.9%, 21.6%, 20.2%, and 19.1% for their stimuli with 1–5 internal nodes, respectively. Because any TSP solution must avoid intersections to be optimal, and because any solution that fails to connect the nodes of the convex hull in order of adjacency must give rise to an intersection, any solution that follows the rule of avoiding intersections must visit the nodes of the convex hull in order (Flood, 1956; Quintas & Supnick, 1965). Thus, number of potential intersections might have contributed to the results of Ormerod and Chronicle (1999).

Fifth, measures of path complexity also appear to vary systematically. Path complexity indexes the extent to which configurations made use of  $k$  nearest neighbors in their construction (Vickers et al., 2004). For each edge in the

configuration, numbers (1 to  $k$ ) are assigned, according to whether the nodes of that link are connected to the nearest (1), second nearest (2), or  $k$ th nearest node. The sum of these numbers is divided by the number of edges to give a measure of path complexity. The lower the path complexity, the more the configuration tends to connect each node to its nearer neighbors. For their stimuli, with 9 to 5 convex hull nodes, respectively, we estimate that path complexity takes values of 3.0, 3.3, 4.2, 5.1, and 6.0. Thus, the data of Ormerod and Chronicle (1999) could be just as consistent with a locally based process of linking near neighbors as they appear to be with a globally based process of using the convex hull as an initial perceptual framework.

Finally, judging from the minuscule standard errors in the two graphs presented in their Figure 2, Ormerod and Chronicle (1999), like MacGregor and Ormerod (1996), do not appear to have found any appreciable individual differences in participants’ responses. As has been argued by Vickers et al. (2001) and by Vickers, Bovet, Lee, and



**Figure 2.** Examples of the stimuli presented to participants. Panels A(i)–A(iii) show examples of stimuli with a medium deviation from the benchmark and a medium number of nodes linked to their nearest neighbors. These panels show instances of stimuli with a low, a medium, and a high number of interior nodes, respectively. Panels B(i)–B(iii) show examples of stimuli with a medium number of interior nodes and a medium number of nodes linked to their nearest neighbors. These panels show instances of stimuli with a low, a medium, and a high medium deviation from the benchmark, respectively. Panels C(i)–C(iii) show examples of stimuli with a medium number of interior nodes and a medium deviation from the benchmark. These panels show instances of stimuli with a low, a medium, and a high number of nodes linked to their nearest neighbors, respectively.

Hughes (2003), such a result is almost certainly due to the relatively sparse and highly constrained arrays employed by them (see also Vickers, Lee, Dry, & Hughes, 2003; Vickers et al., 2004). Because reliable individual differences appear when the problem arrays are less constrained and are more densely populated with nodes, it is possible that such arrays might also give rise to a more complex pattern of ratings of figural goodness than do those obtained by Ormerod and Chronicle (1999).

In view of these difficulties, the following experiment was undertaken to differentiate between the above factors potentially underlying judgments of the figural goodness of TSP solutions and to arbitrate between alternative hypotheses concerning the relationship between these factors and the perceived goodness or aesthetic appeal of such structures. The principal hypotheses, with which we started, were that the figural goodness or aesthetic appeal of solution pathways is based on (1) the optimality of the solution, (2) the number of nodes on the convex hull (or alternatively, the number of internal nodes), or (3) the number of nodes that were linked to their nearest neighbor in a solution. These variables were manipulated in a factorial design that restricted the regularity of the figures (as measured by the standard deviation in their edge lengths) to a very narrow range, excluded actual intersections, and held the number of potential intersections constant. However, the design of the experiment also permitted an investigation of the role of the degree of convexity, circularity, and path complexity and allowed effects of the degree of rotational symmetry (as indexed by the standard deviation in the lengths of the convex hull edges) to be assessed. This was achieved by generating large numbers of random arrays and selecting from them those configurations with the desired characteristics. This method provides a strong contrast with that used by Ormerod and Chronicle (1999) to generate their configurations, which consisted of incremental transformations of one initial, highly regular array. In addition, the use of more complex configurations avoided ceiling effects and allowed for the possible emergence of individual differences.

## METHOD

In this experiment, strict rotational symmetry was excluded by employing randomly generated solution pathways linking well-populated arrays of randomly distributed dots, while approximate rotational symmetry was quantified by the standard deviation in the lengths of the convex hull edges and was allowed to vary randomly. The regularity of the figures, as measured by the standard deviation in the link lengths, was restricted to a very narrow range and was not allowed to vary systematically. There were no actual intersections, and the number of potential intersections was held constant, whereas the number of convex hull nodes, the optimality of solutions, and the number of nodes linked to their nearest neighbors that were incorporated in these solutions were varied in a factorial design. In addition, degree of convexity was allowed to vary randomly, as was the circularity and path complexity of the configurations.

### Stimulus Generation

In order to obtain expectations for the number of nodes on the convex hull, the number of internal node intersections,<sup>2</sup> and the stan-

dard deviation of the internode distance, 20,000 random 25-node arrays were generated within the unit square. For these arrays, the mean number of convex hull nodes was 8.35 ( $SD = 1.41$ ), the mean percentage of internal node intersections was 23.3%, and the mean standard deviation in the internode distance was 0.245.

With these statistics as a basis, we generated a number of random 25-node arrays. Of these, we selected a number that had a low (13), medium (16 or 17), or high (20) number of internal nodes. This selection was also constrained to have a constant proportion (23.0%–23.6%) of internal node intersections and a constant (0.24–0.25) standard deviation in the internode distance. A further subset of these arrays was then chosen, in which the number of internal nodes was visually indisputable. In all, 18 different arrays were chosen: 6 from each of the three internal nodes groups. In addition, 1 from each group was chosen for practice.

For each of these 18 arrays, a benchmark (putatively optimal) closed TSP solution was calculated, using a simulated annealing heuristic (Reinelt, 1994). Random solutions were then generated for each array that had a low (0%–5%), medium (10%–15%), or high (20%–25%) difference from the benchmark solution. (The 0%–25% range of deviations from the benchmark was chosen because it was representative of the range found in the solutions produced by the participants in several of our previous TSP studies.)

For each of these random solutions, we noted the number of nodes that were linked to their nearest neighbor in the solution path. The lowest and the highest numbers of nodes connected to their nearest neighbor that could be found in solutions corresponding to each of the nine difference from benchmark and interior node group combinations were 16 and 23, respectively. Accordingly, we determined that the three nearest neighbor linked groups should be low (15–16), medium (19–20), and high (23–24).

Thus, random solutions were generated for six problem instances in each of the three internal nodes, three differences-from-benchmark, and three nearest-neighbor-linked groups. These solutions were selected to have no path crossings (and as a result, connected the convex hull nodes in sequence). More than one solution was generated for each instance in order that unrealistic tours (ones that incorporated obvious, zigzag backtracking) could be eliminated. However, no other restriction was placed on the solutions. The number of attempts required to generate a single acceptable solution for a given problem instance in a given group ranged from 100 to 1,000,000.

### Stimuli

The participants rated the perceived goodness of 162 configurations, each consisting of complete tours (Hamiltonian circuits) of 25 randomly distributed nodes. The 162 configurations were made up of six instances drawn from each of three (high, medium, and low) internal nodes, differences-from-benchmark, and nearest-neighbor-linked groups (i.e., of 27 level  $\times$  group combinations). The nodes were represented by black circular dots, 1.5 mm in diameter, connected by black links on a light gray background and were displayed within a 14  $\times$  14 cm square, with no border, at the center of a computer screen. Examples of the stimuli are shown in Figure 2. The configurations were viewed from a distance of about 50–60 cm.

### Design

Each participant was presented with the 162 configurations in a different order that was random, but subject to the constraint that all 27 combinations of number of internal nodes, level of optimality, and number of nodes linked to their nearest neighbors were presented once, before being repeated.

### Procedure

Following a practice run of 12 configurations, in which the upper and lower values of each of the three factors were represented, the participants carried out the test run, at their own pace, in one unbroken series. Each instance was initiated by the participant by clicking on a "Proceed to Next Test" button. This was followed by a 1-sec "Get

Ready” cue, succeeded by a 1-sec attention cue, after which the instance and the response boxes were shown on the screen simultaneously.

As has been noted by van der Helm and Leeuwenberg (2004), *figural goodness* is an intuitive Gestalt notion that has no formal definition (see also Olivers, Chater, & Watson, 2004). Because this notion can be assumed to be familiar only to those who know about research into gestalt properties in visual perception, we sought to express the same notion in more widely understandable terms. On the basis that studies of the properties considered *good* by naive participants have used such terms as *most pleasing* and *liked* (Mowatt, 1940), the participants were asked to give a “rating of the aesthetic appeal of the figure, of how attractive they found it as an abstract configuration.” This wording corresponded closely to that employed by Vickers et al. (2001), who instructed participants in their Gestalt group to draw a pathway through the nodes that “looked most natural, attractive, or aesthetically pleasing.” The figures were not described as possible TSP solutions, and no mention was made of path length, optimality, or any other characteristic of the configurations. The rating process involved clicking on one of six response boxes, located along the bottom of the screen and containing both a number and a description. From left to right, these were the following: 0, *very poor*; 1, *poor*; 2, *fairly poor*; 3, *fairly good*; 4, *good*; and 5, *very good*. Each configuration remained until the rating was registered. The entire series of practice and test judgments required around 20–30 min to complete.

### Participants

Forty student volunteers served as participants.

## RESULTS

### Group Analysis

Figure 3 summarizes the relationship between the aesthetic appeal ratings, averaged over participants, and the 27 different combinations of nearest neighbors, convex

hull points, and difference from optimality theoretical measures. The  $3 \times 3$  grid of panels corresponds to the three levels of nearest neighbors (in rows) and convex hull points (in columns). Each panel shows, as a bold line, the mean aesthetic rating averaged across all subjects, with one standard error in each direction, as a function of the level of difference from optimality. Each panel also shows, in gray, the mean values for the other panels, to allow for direct visual comparison.

Figure 3 suggests that changes in the difference from benchmark affect aesthetic appeal, with those configurations that are near-optimal being judged as more appealing. There is also a suggestion of a consistent, but much less pronounced, pattern in relation to the number of convex hull points, with mean aesthetic appeal being greatest for those configurations with the most points on the convex hull.

Because this experiment was designed as a factorial study, our initial analysis was restricted to considering group performance, as summarized in Figure 3. This analysis, however, makes the implicit assumption that there are no meaningful individual differences between participants.<sup>3</sup> If this assumption is incorrect, as was explained long ago by Estes (1956; see also Estes & Maddox, 2005; Lee & Webb, 2005), the averaging of performance measures used to calculate the group data can seriously distort the way those data relate to the underlying psychological processes they are supposed to measure. As it turns out, we believe that there is evidence of significant and meaningful individual differences in the data, and so the majority of our analysis is conducted at the level of individuals.

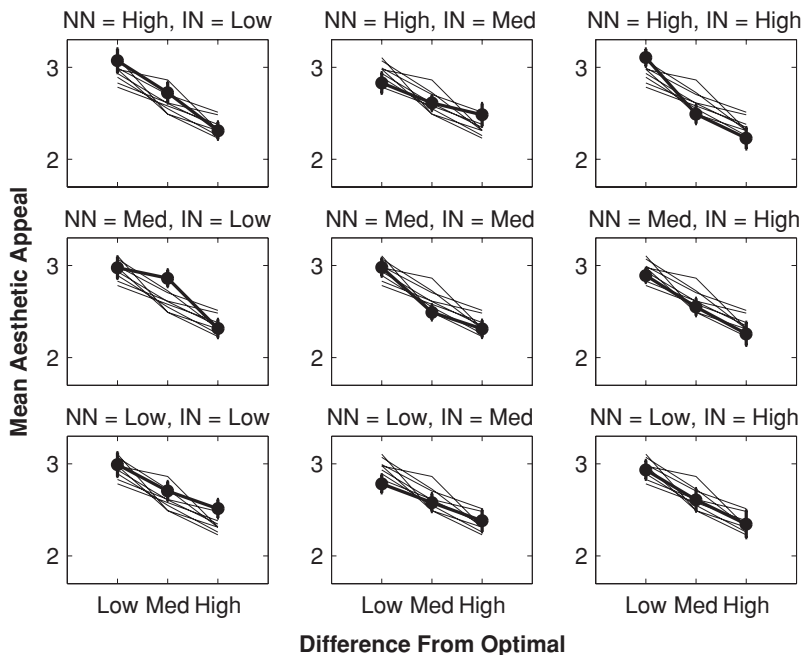


Figure 3. The relationship between the aesthetic appeal ratings and the 27 different combinations of nearest neighbors (NN), number of internal nodes (IN), and difference from optimal theoretical measures.

### Individual Analysis

To examine the reliability of individual differences across the 27 problem types, Pearson correlations were calculated between the average of each participant's ratings for the six problem instances of each type and the average rating for each of the other problem types (i.e., a total of 351 correlations). Of these 351 correlations, there were 224 positive correlations that were significant at the .05 level or better and only 3 significant negative correlations ( $N = 40$  in all cases). The average of the 351 correlations was .41. With 12,990 degrees of freedom, the probability of obtaining an average correlation of this size by chance is minuscule.

Therefore, for each participant, arithmetic mean figural attractiveness ratings were calculated over the six instances of each of the 27 internal nodes, differences-from-benchmark, and nearest-neighbor-linked groups of configurations. Correlations were then calculated, for each participant, between figural attractiveness and each one of a number of stimulus variables.

Several of these variables had no strong theoretical motivation and proved to have little or no predictive value. These included (1) the area inside the convex hull, (2) the total length of the convex hull, (3) the average edge length of the convex hull, and (4) rotational symmetry as indexed by the standard deviation in the edge length of the convex hull. For these variables, arithmetic means (over participants) of the absolute (unsigned) values of their Pearson correlations with figural attractiveness were .12, .34, .33, and .34, respectively.

A number of variables had a stronger theoretical motivation but provided only weak to moderate predictions of attractiveness ratings. Variables in this category included (5) the number of interior nodes, (6) the number of nodes that were linked to their nearest neighbor, and (7) regularity, as indexed by the standard deviation in the lengths of the configuration edges. For these variables, arithmetic means (over participants) of the absolute values of their Pearson correlations with figural attractiveness were .34, .14, and .45, respectively. Of these, only the measure of regularity made a moderate contribution to the prediction of attractiveness ratings.

Finally, four variables possessed a strong theoretical motivation and made a moderate to strong contribution to the prediction of attractiveness ratings. These variables were (8) the percentage by which the solution length exceeded the benchmark (putatively optimal) solution, (9) the circularity of the configurations, (10) convexity, and (11) path complexity. For these variables, arithmetic means (over participants) of the absolute values of their Pearson correlations with figural attractiveness were .58, .66, .47, and .62, respectively.

Table 1 summarizes the intercorrelations among each of the last four variables and their correlations with rated attractiveness. For rated attractiveness, two sets of correlations are shown: (1) correlations between mean stimulus measures (averaged over the six instances of each of the 27 stimulus types) and mean attractiveness ratings for each stimulus type (averaged over all participants and over each of the stimulus types) and (2) averages of the absolute values of the correlations between mean stimulus measures (averaged for each of the 27 stimulus types) and mean attractiveness ratings for each participant (averaged for each of the 27 stimulus types).

Figure 4 shows raw correlations for individual participants between figural attractiveness ratings and the four most successful measures (circularity, path complexity, percentage above benchmark, and convexity). The most successful predictive measure overall turned out to be circularity. Therefore, the participants have been ordered according to the strength of the correlation (over the 27 configuration types) between their ratings of figural attractiveness and the mean circularity measures for the corresponding configuration types. (Hence, the correlations for circularity necessarily show a monotonic decrease from left to right in Figure 4.)

The most striking feature of Figure 4 is that the 40 participants naturally fall into two distinct subgroups. Participants 1–30 show positive correlations between figural attractiveness and circularity, whereas Participants 31–40 show negative correlations between these two measures. With the exception of two points, the correlations for the other three measures follow a similar pattern. In recognition of this pattern of individual differences, the four

**Table 1**  
Intercorrelations Among Measures of Circularity, Optimality, Path Complexity, and Convexity and Correlations Between These Measures and Rated Attractiveness of the 27 Stimulus Types

	Circularity	Optimality	Path Complexity	Convexity
Attractiveness (a)	.84	-.92	-.90	.37
Attractiveness (b)	.66	.58	.62	.47
Circularity		-.82	-.93	.75
Optimality			.93	-.25
Path complexity				-.52

Note—In the case of rated attractiveness, two sets of correlations are shown: (a) correlations between mean stimulus measures for each of the 27 stimulus types and group mean attractiveness ratings of each stimulus type (i.e., averaged over all participants) and (b) (unsigned) averages of the absolute values of the 40 correlations between mean stimulus measures for each of the 27 stimulus types and mean attractiveness ratings by each individual participant for each stimulus type.

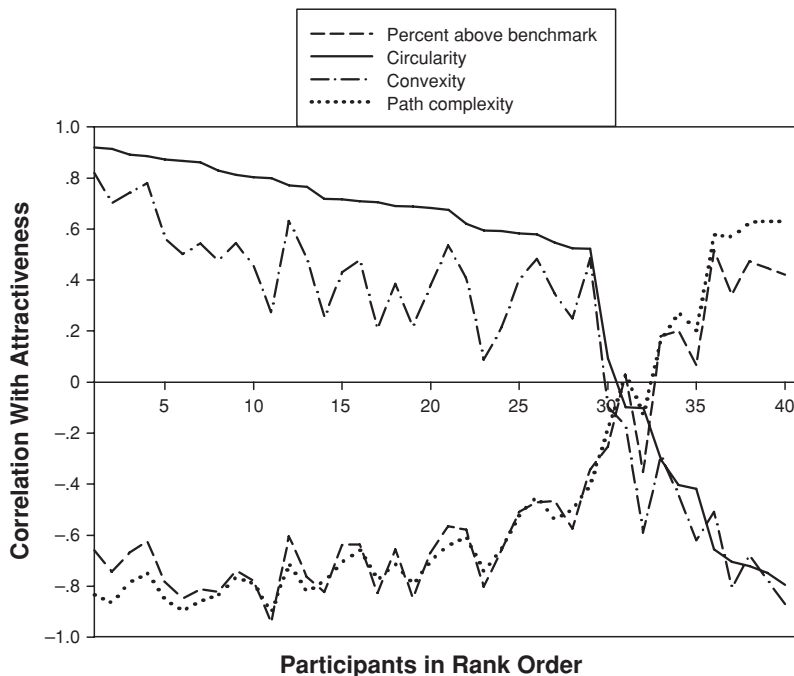


Figure 4. Signed values of the 40 correlations between attractiveness ratings by individual participants (averaged over instances of each of the 27 stimulus types) and mean values of the four most predictive stimulus configuration measures (also averaged over instances of each of the 27 stimulus types). The participants have been rank ordered by the (signed) value of the correlation between their ratings and measures of circularity. Although values on the abscissa represent discrete rank order, the corresponding points on the ordinate have been joined by continuous lines to make the trends clearer.

most predictive measures were examined, for different partitionings of the participants, to see which measure produced the highest positive and negative average correlations for a given partitioning. The measure that was most successful in this respect was circularity, with participants partitioned into Subgroups 1–30 (average  $r = .94$ ) and 31–40 (average  $r = -.80$ ). For the same partitioning, path complexity produced average correlations of  $r = -.93$  (Participants 1–30) and  $r = .60$  (Participants 31–40). Convexity produced correlations of  $r = .58$  (Participants 1–30) and  $r = -.90$  (Participants 31–40), respectively. By exchanging the rank position of 2 participants (i.e., by ranking the participants according to the correlation between their attractiveness ratings and solution optimality), percentage above benchmark produced average correlations of  $r = -.89$  and  $.43$ , respectively.

## DISCUSSION

The results presented above do not lend themselves to easy generalizations over participants. Nevertheless, a number of general conclusions can be drawn. These concern the importance of individual differences in judgments, the possibility of polar differences in perceptual style, and the relative importance of different stimulus

measures for judgments of figural goodness or attractiveness. We shall discuss each point in turn.

### Individual Differences

The first point is that individual differences in ratings of figural attractiveness in the present study were highly reliable, with ratings for any one solution type being positively and significantly correlated with ratings for the great majority of the other solution types. These data are different from those in Ormerod and Chronicle's (1999) Experiment 1, which showed negligible variation across participants in their ratings of figural goodness. On the other hand, the present result is in agreement with other findings of consistent individual differences across problem instances and across different types of problems, as well as reliable correlations with psychometric measures of intelligence (Vickers, Bovet, et al., 2003; Vickers et al., 2001; Vickers, Lee, et al., 2003; Vickers et al., 2004).

It is not possible to explain the divergence, with respect to individual differences, between our results, either here or in our other studies, and those of MacGregor and colleagues (e.g., MacGregor & Ormerod, 1996; MacGregor, Ormerod, & Chronicle, 1999) in terms of the provision or otherwise of feedback or knowledge of results. It is true that the experiments of Vickers, Bovet, et al. (2003) and

of Vickers, Lee, et al. (2003) provided trial-by-trial feedback by way of correct solutions. However, neither the present experiment nor the four experiments reported by Vickers et al. (2001) and by Vickers et al. (2004) provided any feedback or knowledge of results, but all showed reliable individual differences.

It is also not possible to account for this divergence in terms of a difference in the opportunities for learning in our experiments and in those of MacGregor and colleagues. There is nothing to be learned in the present experiment. Meanwhile, Experiment 1 in Vickers et al. (2001) involved six different problem instances, whereas their Experiment 2 involved only one. Similarly, Experiment 1 in Vickers et al. (2004) used five different problem instances, whereas their Experiment 2 involved only one instance of each of three different types of problems. Although the participants in these experiments were presented with fewer problem instances and had *less* opportunity for learning than were those in the experiments of MacGregor and colleagues, consistent individual differences still emerged. As was suggested by Vickers, Bovet, et al. (2003), the most plausible explanation for the divergence between the two sets of results stems from the fact that the stimuli used by MacGregor and colleagues contain fewer nodes and were generated in a highly constrained way, so that the problems were simpler and variations in performance were limited by ceiling effects. Similarly, the discrepancy between the present results and those in Ormerod and Chronicle (1999) appear to be due to several potentially confounding factors in the latter study that combined to produce responses that were overdetermined by correlated information.

### Differences in Perceptual Style

The second point, which is clearly made in Figure 2, is that individual differences in this type of judgment consist not only of reliable variations in the extent to which certain stimulus variables appear to determine ratings of figural attractiveness, but also of differences in the nature of the relation (positive or negative) between these variables and attractiveness ratings. With the exception of one or two borderline individuals, the participants in this experiment fell into two quite distinct groups: a larger group (Participants 1–30), for whom circularity and convexity were positively related to attractiveness ratings, whereas path complexity and percentage above the benchmark were negatively correlated, and a smaller group (Participants 31–40), for whom such variables were related to rated attractiveness in the opposite way. Although we did not anticipate this outcome, in hindsight, we should have been forewarned, because the task of producing natural, attractive, or aesthetically pleasing configurations, employed by Vickers et al. (2001) with their gestalt group of participants, resulted in a similar wide spectrum of approaches that “ranged from pathway minimization . . . up to one or two that seemed designed to maximize the pathway length” (p. 40).

At this stage, we can only speculate about the explanation for this polar opposition between the two groups. There are a number of other oppositions between individual differences in *perceptual style* that have been investigated under the rubrics of field dependence/independence (Streibel & Ebenholtz, 1982), local versus global processing (Hogeboom & van Leeuwen, 1997), and analytic versus holistic processing (Cooper, 1976; Just & Carpenter, 1985). The contrast that seems to be captured in these oppositions is one between a parallel process, in which perception is determined in an immediate way by the overall configuration of a stimulus, and a more deliberate, serial process, in which attention is focused on figural elements. This contrast characterizes quite well the opposition between the global, convex hull approach to understanding human TSP performance developed by MacGregor and his colleagues (see, e.g., MacGregor & Ormerod, 1996; MacGregor, Chronicle, & Ormerod, 2004) and the locally focused approach proposed by us (Vickers, Bovet, et al., 2003; Vickers, Lee, et al., 2003), which is based on relational information of the kind provided by nearest neighbors.

However, although this polar opposition might conceivably have some relevance for the understanding of TSP performance, it does not seem to map onto the opposition shown in the present data. The reason is that individual participants appear to use both global and local features of the configurations in the same way, but both global and local features seem to play opposite roles in the judgments of the two groups of participants. Thus, our results echo those of Jacobsen (2004), who found that although intraindividual judgments of aesthetic attraction were consistent, their relationships to stimulus properties could not be adequately represented by a model based on average responses.

### Factors Influencing Figural Attractiveness

Despite the problem for any model based on averages, it is still possible to make some general inferences from the data regarding the relative importance of different stimulus features in determining judgments of attractiveness or figural goodness. For example, among the theoretically motivated variables that turned out to have little or no predictive value, the number of nodes that were connected to their nearest neighbor produced the clearest result. The arithmetic mean (over the 40 participants) of the absolute values of the Pearson correlations between this variable and rated attractiveness was very low ( $r = .14$ ); over the 162 individual configurations, this correlation dropped to .07. This suggests that figural attractiveness is not based on the extent to which a configuration incorporates nearest neighbor links.

On the other hand, the optimality of solution pathways in this study was determined largely by the extent to which such pathways incorporated higher order nearest neighbors. This is reflected in the high correlations between optimality and path complexity ( $r = .93$ ) and between judged attractiveness and path complexity (average



absolute  $r$  over 27 stimulus groups = .62). This suggests that the attractiveness of TSP solutions may be influenced (positively or negatively for different participant groups) by the extent to which the solutions approach optimality and have lower values of path complexity.

Of the other measures with a strong theoretical motivation, the number of internal nodes (or conversely, the number of nodes on the convex hull) was not very successful. The arithmetic mean (over the 40 participants) of the absolute values of the Pearson correlations between this variable and rated attractiveness was low ( $r = .34$ ). It appears that participants show consistent preferences between configurations for which the number of internal nodes is the same. As was suggested earlier, the number of nodes on the convex hull may not provide a useful predictive measure of TSP performance either, once the total number of nodes exceeds about 50. It is also the case that a measure of convexity might be expected to prove a more sensitive and more successful predictor of attractiveness ratings, as was the case in this experiment (average absolute  $r$  over 27 stimulus groups = .47). Indeed, although convexity was much less successful than percentage above benchmark, circularity, or path complexity in predicting attractiveness ratings by the larger group of participants, it did provide the strongest prediction of attractiveness by the smaller group (Participants 31–40). At the same time, convexity was strongly correlated with circularity ( $r = .75$ ), so that it is possible that this accounts for some of its predictive value.

The measure that provided the strongest prediction of figural attractiveness overall for both groups was circularity (average absolute  $r$  over 27 stimulus groups = .66). For the larger group of participants, this could be explained as being due to a preference for a structure that is optimal in the sense of enclosing a given area (or volume) with the shortest perimeter (or the smallest surface area in the case of a three-dimensional object). This optimization process may be subject to particular constraints in a given situation, which may result in more complex structures than circles or spheres. However, such structures are ubiquitous in nature as well as in human constructions and are generally regarded as both aesthetically and practically satisfying (Hildebrandt & Tromba, 1996).

This leaves the problem of why the smaller group responded in an opposite way to circularity, as well as to the other three most predictive measures. The most plausible explanation we can suggest is that both groups of participants have an aesthetic preference for structures that are, in some sense, optimal. However, no mention was made of optimality in this experiment, nor was there any indication of what might constitute optimality. Besides natural and artificial structures that enclose a given area (or volume) within the shortest perimeter (or smallest surface area), there are also natural structures that are optimal in exposing a given area (or volume) to the greatest possibility of interchange with the immediate environment of that area (or volume). Examples include lung tissue and numerous branching structures, such as coral or veins

and arteries that have fractal characteristics (Mandelbrot, 1983). Such structures also appear to have an aesthetic attraction, and it is possible that the smaller group of participants might have judged the configurations as more or less optimal in this sense.

Although the present results are equivocal with respect to what participants regard as *optimal* structure, it would be possible to investigate whether aesthetic judgments of a wide range of different two-dimensional configurations and three-dimensional forms exhibit a similar pattern of maximizing or minimizing contour or surface for a given enclosed area or volume. It would also be possible to examine the extent to which such preferences might be influenced by cognitive factors, such as participants' understanding of what such configurations or forms might represent and of what a given structure might be designed to optimize.

Meanwhile, there is no information in the present experiment that supports a strong link between the individual differences found here and other differences in perceptual style or personality. However, if the account above in terms of preference for different types of optimal structure is assumed, we might interpret this as a difference in preference for forms that minimize or maximize the boundaries at which exchanges with the surrounding environment can take place. If so, we might go on to speculate that the pattern of individual differences found in the present experiment could be linked to such personality traits as openness (Gignac, Stough, & Loukomitis, 2004) and preference for complexity (Berlyne, 1974; Marković & Gvozdenović, 2001). However, only further research can decide whether such speculation may prove fruitful.

## CONCLUSIONS

The present study identifies several confounding factors that may have contributed to the conclusion by Ormerod and Chronicle (1999) that the identification of the convex hull is important in determining an overall gestalt and that the greater the number of solution nodes that fall on the convex hull, the more attractive the solution will be judged to be. Results from the present study show that the number of convex hull nodes was only a weak predictor (either positively or negatively) of attractiveness ratings. Taken in conjunction with Vickers, Lee, et al.'s (2003) finding that the greater the number of points on the convex hull of randomly generated arrays, the *poorer* were participants' solutions, the present results argue against the influence of the convex hull as an important determinant of perceptual organization, of the optimality of participants' solutions, or of the perceived goodness or attractiveness of TSP solutions.

On the other hand, the present results provide qualified support for Ormerod and Chronicle's (1999) finding that participants' ratings of the goodness of TSP solutions were correlated with their relative optimality. The important qualification is that what is defined as optimal by the experimenter (in this case, minimizing total path length)

may not be what the participants regard as optimal. Instead, it appears that participants are divided as to whether the most attractive configuration is that which encloses a given area within a perimeter of minimum or of maximum length (i.e., one that is most or least circular).

The results did not support the view that solutions would be judged attractive to the extent that they incorporated nearest neighbors. On the other hand, the results were consistent with the view that good organization was determined by path complexity (i.e., the extent to which a solution incorporated near neighbors of the lowest possible order). The results were also consistent with the hypothesis that judged attractiveness is influenced by convexity, although to a lesser extent. Thus, the results are consistent with the possibility that configurations are judged attractive to the extent that they are optimal and incorporate spontaneous, locally focused, perceptual organizing principles based on nearest neighbors. Similarly, the results do not rule out the possibility that attractiveness is based (albeit less strongly) on a globally focused measure. If so, however, the most promising measure is that of convexity, rather than the number of nodes on the convex hull. To test such hypotheses, future work would need to achieve the independent manipulation of path complexity, percentage above benchmark, circularity, and convexity.

In contrast to the study of Ormerod and Chronicle (1999), the present experiment showed consistent individual differences between participants. Each of the configuration types was rated by the participants in much the same way as each of the other configuration types. However, the attractiveness ratings for one group of participants appeared to be related to stimulus characteristics in a way that was opposite to that of the other group. Although both groups may be construed as preferring structures that are optimal (albeit in opposing ways), it seems clear that future research in this area cannot afford to focus exclusively on group performance measures.

## REFERENCES

- ATTNEAVE, F. (1982). Prägnanz and soap bubble systems: A theoretical exploration. In J. Beck (Ed.), *Organization and representation in perception* (pp. 11-29). Hillsdale, NJ: Erlbaum.
- BERLYNE, D. E. (1974). Novelty, complexity, and interestingness. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics* (pp. 175-180). Washington, DC: Hemisphere.
- BERTAMINI, M., & CROUCHER, C. J. (2003). The shape of holes. *Cognition*, **87**, 33-54.
- CHI, M. T. H., GLASER, R., & FARR, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- COOPER, L. A. (1976). Individual differences in visual comparison processes. *Perception & Psychophysics*, **19**, 433-444.
- DUNCAN, C. P. (1959). Recent research on human problem solving. *Psychological Bulletin*, **56**, 397-429.
- ESTES, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, **53**, 134-140.
- ESTES, W. K., & MADDOX, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, **12**, 403-408.
- FLOOD, M. M. (1956). The traveling salesman problem. *Operations Research*, **4**, 61-75.
- GIGNAC, G. E., STOUGH, C., & LOUKOMITIS, S. (2004). Openness, intelligence, and self-report intelligence. *Intelligence*, **32**, 133-143.
- GRAHAM, S. M., JOSHI, A., & PIZLO, Z. (2000). The traveling salesman problem: A hierarchical model. *Memory & Cognition*, **28**, 1191-1204.
- HILDEBRANDT, S., & TROMBA, A. (1996). *The parsimonious universe: Shape and form in the natural world*. New York: Springer.
- HOGEBOM, M., & VAN LEEUWEN, C. (1997). Visual search strategy and perceptual organization covary with individual preference and structural complexity. *Acta Psychologica*, **95**, 141-164.
- JACOBSEN, T. (2004). Individual and group modelling of aesthetic judgment strategies. *British Journal of Psychology*, **95**, 41-56.
- JUST, M. A., & CARPENTER, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, **92**, 137-192.
- KANIZSA, G., & GERBINO, W. (1976). Convexity and symmetry in figure-ground organization. In M. Henle (Ed.), *Vision and artifact* (pp. 25-32). New York: Springer.
- LAWLER, E. L., LENSTRA, J. K., RINNOOY KAN, A. H. G., & SHMOYS, D. B. (1985). *The traveling salesman problem: A guided tour of combinatorial optimization*. Chichester, U.K.: Wiley.
- LEE, M. D., & VICKERS, D. (2000). The importance of the convex hull for human performance on the traveling salesman problem: A comment on MacGregor and Ormerod (1996). *Perception & Psychophysics*, **62**, 226-228.
- LEE, M. D., & WEBB, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, **12**, 605-621.
- LIU, Z., JACOBS, D. W., & BASRI, R. (1999). The role of convexity in perceptual completion: Beyond good continuation. *Vision Research*, **39**, 4244-4257.
- MACGREGOR, J. N., CHRONICLE, E. P., & ORMEROD, T. C. (2004). Convex hull or crossing avoidance? Solution heuristics in the traveling salesperson problem. *Memory & Cognition*, **32**, 260-270.
- MACGREGOR, J. N., & ORMEROD, T. [C.] (1996). Human performance on the traveling salesman problem. *Perception & Psychophysics*, **58**, 527-539.
- MACGREGOR, J. N., & ORMEROD, T. C. (2000). Evaluating the importance of the convex hull in solving the Euclidean version of the traveling salesperson problem: Reply to Lee and Vickers (2000). *Perception & Psychophysics*, **62**, 1501-1503.
- MACGREGOR, J. N., ORMEROD, T. C., & CHRONICLE, E. P. (1999). Spatial and contextual factors in human performance on the travelling salesperson 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**, 1183-1190.
- MANDELBROT, B. B. (1983). *The fractal geometry of nature*. New York: Freeman.
- MARKOVIĆ, S., & GVOZDENOVIĆ, V. (2001). Symmetry, complexity and perceptual economy: Effects of minimum and maximum simplicity conditions. *Visual Cognition*, **8**, 305-327.
- METZGER, F. (1953). *Gesetze des Sehens*. Frankfurt-am-Main: Waldemar Kramer.
- MOWATT, M. H. (1940). Configurational properties considered "good" by naive subjects. *American Journal of Psychology*, **53**, 46-69.
- OLIVERS, C. N. L., CHATER, N., & WATSON, D. G. (2004). Holography does not account for goodness: A critique of van der Helm and Leeuwenberg (1996). *Psychological Review*, **111**, 242-260.
- ORMEROD, T. C., & CHRONICLE, E. P. (1999). Global perceptual processing in problem solving: The case of the traveling salesperson. *Perception & Psychophysics*, **61**, 1227-1238.
- O'ROURKE, J. (1993). *Computational geometry in C*. Cambridge: Cambridge University Press.
- POLIVANOVA, N. I. (1974). [Functional and structural aspects of the visual components of intuition in problem solving]. *Voprosy Psikhologii*, **4**, 41-51.
- QUINTAS, L. V., & SUPNICK, F. (1965). On some properties of shortest Hamiltonian circuits. *American Mathematical Monthly*, **72**, 977-980.
- REINELT, G. (1994). *The traveling salesman: Computational solutions for TSP applications* (Lecture Notes in Computer Science, Vol. 840). Berlin: Springer.
- SELKIRK, K. E. (1982). *Pattern and place*. Cambridge: Cambridge University Press.
- STEIN, D. L. (1989). Spin glasses. *Scientific American*, **261**, 36-42.

- STREIBEL, M. J., & EBENHOLTZ, S. M. (1982). Construct validity of perceptual style: Role of stimulus size in the embedded-figures test and the rod-and-frame test. *Perception & Psychophysics*, **31**, 128-138.
- VAN DER HELM, P. A., & LEEUWENBERG, E. L. J. (2004). Holographic goodness is not that bad: Reply to Olivers, Chater, and Watson (2004). *Psychological Review*, **111**, 261-273.
- 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**, 871-886.
- VICKERS, D., BUTAVICIUS, M., 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 the number of potential intersections in performance on visually presented traveling salesperson problems. *Memory & Cognition*, **31**, 1094-1104.
- 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 & Individual Differences*, **36**, 1059-1071.
- ZUNIC, J., & ROSIN, P. L. (2004). A new convexity measure for polygons. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **26**, 923-934.

#### NOTES

1. It might be useful to reformulate the convex hull hypothesis in terms of convexity, rather than simply in terms of the number of convex

hull nodes. Indeed, Lee and Vickers (2000) have shown that, because the number of nodes on the convex hull exhibits asymptotic behavior as the total number of nodes is increased, the potential influence of this factor is limited to arrays with around 50 nodes or fewer (see also MacGregor & Ormerod, 2000; Vickers, Bovet, Lee, & Hughes, 2003).

2. Potential intersections can be counted for all possible pairs of links between all possible pairs of nodes (without any pair of links having a common node). The proportion of pairs of links that result in intersections and involve all possible nodes remains roughly constant (at around 23.2% of all possible pairs of links), irrespective of the number of nodes in an array. Similarly, the number of pairs of links that result in intersections and that involve internal nodes only remains a roughly constant proportion (around 23.3% of all possible links between internal nodes). This proportion remains roughly constant, even when the number of nodes on the convex hull varies. In contrast, the number (and proportion) of intersections involving all possible nodes increases as a function of the number of nodes on the convex hull. For this reason, it was decided to manipulate the number of potential intersections involving internal nodes only, so as to minimize any possible interaction between the proportion of intersections involving internal nodes and the number of nodes on the convex hull.

3. We are indebted to an anonymous referee of an earlier version of this article for encouraging us to examine individual differences.

(Manuscript received February 24, 2004;  
revision accepted for publication March 7, 2005.)