

Perceptual or Analytical Processing? Evidence from Children's and Adult's Performance on the Euclidean Traveling Salesperson Problem

Iris van Rooij,¹ Alissa Schactman,² Helena Kadlec,³ and Ulrike Stege⁴

Abstract

The Euclidean Traveling Salesperson Problem (E-TSP) is a useful task to study how humans optimize when faced with computational intractability. It has been found that humans are capable of finding high-quality solutions for E-TSP in a relatively short time and with seemingly little cognitive effort. This observation has led to two general proposals: The high quality of performance on E-TSP reflects (a) the output of automatic and innate perceptual processes or (b) a fundamentally intelligent approach to the task by humans. An experiment was performed to compare performance of three age groups (7-year-olds, 12-year-olds, and adults). Our findings provide corroborating evidence that high-level performance on E-TSP is attainable with perceptual processing alone. At the same time, the exceptionally high performance of adult participants suggests that they additionally exploit their cognitive-analytic skills to improve upon what their visual systems provide.

Introduction

The notion of optimization is a recurring theme in cognitive science. Traditionally, optimization has been seen as fundamental to human rationality (Edwards, 1954; Simon, 1957; von Neuman & Morgenstern, 1947). Also, some have proposed that, in a sense, every cognitive system instantiates an optimal response to some problem situation (e.g., Anderson, 1990; Marr, 1982) or, at least, that optimization is part and parcel of human intelligence (Vickers, Butavicius, Lee, & Medvedev, 2001). The computation of optimal responses sometimes exceeds the computational resources that humans have available, and thus, human rationality is necessarily bounded (e.g., Oaksford & Chater, 1998; Simon, 1988; 1990). Nonetheless, humans (still) far exceed any computing machine

¹Human-Technology Interaction, Eindhoven, University of Technology Eindhoven, The Netherlands

²Department of Psychology, University of Saskatchewan, Saskatoon, Canada

³Department of Psychology, University of Victoria, Victoria, Canada

⁴Department of Computer Science, University of Victoria, Victoria, Canada

Correspondence concerning this article should be addressed to: Iris van Rooij, Human-Technology Interaction, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, Telephone: +31(0)40 2473163, Facsimile: +31(0)40 2449875, Email: i.v.rooij@tm.tue.nl

in their ability to respond reasonably, flexibly, and adaptively in real-world, complex problem situations. The desire to understand this admirable feat lies at the very foundation of both cognitive science and artificial intelligence (e.g., Haselager, 1997; Pylyshyn, 1987). It is in this context that the study of human performance on optimization problems, such as the Euclidean Traveling Salesperson Problem, is of particular interest.

In the Euclidean Traveling Salesperson Problem (E-TSP), one is given a set of points in the Euclidean plane and the goal is to determine an optimal tour; that is, to find a closed path that visits each point in the set, starting and ending at the same point, such that the total length of the path is minimized. E-TSP has several useful features for the experimental study of human optimization. First, it instantiates a computationally intractable¹ task and, thus, captures something about the complexity of real-world tasks that is not captured in other kinds of problem-solving tasks. Second, the task is constrained enough that it can be brought into the laboratory, allowing for a relatively high degree of experimental control. Third, despite its computational complexity, the formulation of the problem is simple and easy to understand for participants. For these reasons, E-TSP constitutes a useful test-case in the empirical study of human optimization in the face of computational complexity.

Several experimental studies of human performance on E-TSP have been performed to date (Graham, Joshi, & Pizlo, 2000; MacGregor & Ormerod, 1996; MacGregor, Ormerod, & Chronicle, 1999, 2000; Polivanova, 1974; Vickers et al., 2001). In these studies, participants are typically presented with a set of points on paper (or on a computer screen, Graham et al., 2000) and are then instructed to draw the shortest possible tour through the points (see Polivanova, 1974, for an alternative presentation format). It has been found that, when presented in this form, participants are able to find close-to-optimal tours in a relatively short time (Graham et al., 2000; Vickers et al., 2001) and with seemingly little cognitive effort (MacGregor & Ormerod, 1996; MacGregor et al., 1999, 2000; Polivanova, 1974). This finding contrasts with the apparent computational complexity of E-TSP and has led to two general theoretical proposals.

MacGregor and colleagues have proposed that the high quality of human performance on E-TSP reflects the workings of fixed, possibly innate, perceptual mechanisms (MacGregor & Ormerod, 1996; MacGregor et al., 1999, 2000). More specifically, these authors proposed that the task of finding the shortest tour coincides with a natural tendency of the perceptual system to perceive *minimum structure* or *good form* (cf., Koffka, 1935). This idea has found further support in the study by Ormerod and Chronicle (1999), who presented participants with predrawn tours and found that participants rated shorter tours as "better figures." In contrast, Vickers et al. (2001) have proposed that the high quality of human performance reflects intelligent, adaptive, and task-specific cognitive processing by participants. In support of their interpretation, they reported systematic individual differences in task performance that also significantly correlated with general intelligence tests.

The two proposals, by MacGregor and colleagues on the one hand and by Vickers and colleagues on the other, need not be seen as categorically opposing viewpoints. On the contrary, part of an intelligent approach to E-TSP may be to flexibly and adaptively exploit natural perceptual tendencies. Such a combined view is also more consistent with Polivanova's (1974) original proposal that both perceptual and analytical processing contribute to human performance on E-TSP. In this paper, we set out to provide further support for such a two-process view by studying specifically the performance of *children* on E-TSP. We begin by explaining how we think the study of children's performance may provide insight into the bases of adult performance on E-TSP.

Rationale of the Study

In our experiment, we compared the performance of three different age groups. The younger children were around the age of 7; the older children were around the age of 12; and the adults had an average age of about 19 (see [Methods](#) section for more details). On the one hand, because perceptual (Gestalt) organization develops in infancy (Johnson & Bremner, 2000; Slater, 2000; Quinn & Bhatt, 2005), we can assume that perceptual organization of point sets is the same for 7-year-olds, 12-year-olds, and adults. On the other hand, these three age groups differ greatly in level of cognitive-analytic development. According to Piaget's stage theory of child development, at the age of about 7, children make the transition from the preoperation stage to the concrete-operational stage, and at the age of about 12, children make the transition from the concrete-operational stage to the formal operations stage (Ginsberg & Opper, 1979; Inhelder & Piaget, 1958). Although Piaget's stage theory remains controversial, it is generally accepted by developmental psychologists that 7-year-olds, 12-year-olds, and adults differ substantially in terms of cognitive-analytical abilities. In sum, the three groups are comparable with respect to perceptual resources but differ in cognitive-analytic skills that they can bring to bear on the task. Reasonably, task performance of the youngest age group will be driven mostly by perceptual processes. In this paper we will, therefore, work under the assumption that (a) performance of the youngest group provides us with an estimate of the level of performance attainable with little or no cognitive-analytic processing and (b) any observed differences between the age groups reflect the relative contribution of cognitive-analytic processing.

Our experiment can be seen as extending Experiment 1 in Vickers et al. (2001). Vickers et al. compared performance of two groups of participants. The first group, the *Optimization group* (O-group), received the standard E-TSP instruction to draw the shortest possible tour. The second group, the *Gestalt group* (G-group), was instead instructed to draw a tour that "looked most natural, attractive, or aesthetically pleasing" (Vickers et al., 2001, p. 36), what in the Gestalt school of psychology is referred to as "good form" or *Prägnanz* (see any introductory text, e.g., Nairne, Smith, & Lindsay, 2001). Vickers et al.

found that (a) participants in both groups produced optimal tours about equally often; (b) the shortest tour produced by the G-group for a given point set was, on average, not systematically longer than the shortest tour produced by the O-group on the same point set; (c) a sizable proportion of tours produced by the G-group (around 30.5% overall) were identical to tours produced by the O-group; (d) there was larger variability in the lengths of tours produced by the G-group as compared to the O-group; (e) tours produced by the O-group were on average shorter than tours produced by the G-group; and (f) the G-group was on average faster than the O-group in constructing tours.

Findings (a), (b), and (c) support the notion that E-TSP coincides, at least to some extent, with natural perceptual tendencies of the cognitive system (i.e., its propensity for perceiving and producing good forms). Findings (d), (e), and (f) suggest, however, that human performance on E-TSP cannot be understood solely in terms of such task-independent perceptual processing. As discussed by Vickers et al. (2001, p. 40), findings (d) and (e) were largely due to the G-group participants' producing tours of qualitatively different character than the tours produced by the O-group. That is, in addition to tours that seemed to aim at minimizing tour length, the G-group also produced some tours that seemed to aim at constructing spiral-like structures and maximizing tour length. This suggests that very different kinds of tours may all be of a similarly "good form" (e.g., in terms of rotational symmetry and good continuation), and thus perceiving a good form is not sufficient for deriving specifically *short* tours. Finding (f) further suggests that the task of producing specifically short(est) tours requires a process that is more time-consuming than Gestalt perception.

Notably, because of the difference in instructions to the G- and O-group, the findings by Vickers et al. can be seen as consistent with the hypothesis that little or no cognitive-analytical processing occurs in the O-group. Namely, it may be that participants in both groups perceived multiple good forms in any given point set and that the G-group and O-group differed only in that the latter specifically selected the shorter of these tours for production. Such a simple selection mechanism, possibly entailing some search mechanism through a set of candidate tours, can also account for the difference in speed of performance of the two groups. The manipulation of age in our experiment, albeit quasi-experimental, has the advantage that all groups can receive the same instructions, and thus differences between the age groups are not due to differences in task but, most likely, due to differences in cognitive-analytic competence.

To facilitate differentiating child and adult performance we furthermore manipulated the number of points in the point set (5, 10, or 15 points). Many studies manipulate instance size,² and consequently, effects of this manipulation are well-documented for adult participants. The effect of instance size is of particular interest in the broader context of the study of human optimization. A common criticism of many cognitive theories has been that—even though they may successfully model how humans solve relatively small and simple problems—they fail to scale up to problems of real-world size and

complexity (Haselager, 1997; Pylyshyn, 1987). It is known that the computational resources required to solve E-TSP grow extremely fast with instance size (Garey & Johnson, 1979). No known practical algorithm solves E-TSP optimally while avoiding an exhaustive search on the $\frac{(n-1)!}{2}$ possible tours for n points. Note that, even though there exist only 12 possible tours for any given 5-point instance, the number of possible tours for a 15-point instance already exceeds 4×10^{10} . Without some way of limiting the space of candidate solutions, solving E-TSP can be like searching for a needle in a haystack. For this reason, it is of interest to study how human performance on E-TSP scales with instance size. Furthermore, a comparison of the effect of instance size in the three age groups provides us with useful information about the differences and commonalities in the groups' approaches to the task.

Measurement Issues

In order to describe the performance of each age group as precisely as possible, we used several performance measures falling into three general classes: (a) measures of tour quality, (b) between-subjects variability, and (c) violations of optimality rules. We discuss each in turn and review the effects of instance size reported in the literature. [Table 1](#) presents an overview.

Measures of Tour Quality

All psychological studies on E-TSP use some measure of tour quality. Because the task of E-TSP is to minimize tour length, tour quality is reflected by the degree to which participants succeed in producing short(est) tours. A first possible measure of quality, then, is simply the length of the tours produced by participants. The problem with this measure, however, is that different point sets (possibly of different sizes) may afford tours of very different lengths simply because of the particular locations of the points in the Euclidean plane. To circumvent this problem, MacGregor and Ormerod (1996) used *standardized tour lengths* instead, defined as:

$$z_L = \frac{L_{obs} - \mu_{rand}}{\sigma_{rand}} \quad (1)$$

where L_{obs} denotes the length of the observed tour (the participant's response), and μ_{rand} and σ_{rand} denote the mean and the standard deviation of the lengths of all possible tours for the instance in question (estimated by a random sample of tours when the number of possible tours is too large). MacGregor and Ormerod estimated μ_{rand} and σ_{rand} from random samples of 10,000 tours per instance and found that participants' absolute z_L scores were larger for their 20-point instances than for their 10-point instances (see Tables 1 and 2 in MacGregor & Ormerod, 1996).

Table 1

Summary of Assessment Measures of Human Performance on the E-TSP.

Measure	Definition	Relevant Empirical Findings
<i>Measures of Tour Quality</i>		
z_L	$\frac{L_{obs} - \mu_{rand}}{\sigma_{rand}}$	Increases with instance size n (MacGregor & Ormerod, 1996)
d_N	$\frac{L_{obs} - L_{opt}}{L_{opt}}$	Increases (almost linearly) with n (Graham et al., 2000; MacGregor & Ormerod, 1996; MacGregor et al., 1999)
z_d	$\frac{L_{obs} - L_{opt}}{\sigma_{rand}}$	Nonlinear trend (Vickers et al., 2001).
p_{opt}	Proportion of instances for which the optimal tour was found	Decreases (slowly) with n (Graham et al., 2000; MacGregor et al., 1999, 2000; Vickers et al., 2001)
<i>Measures of Between-Subject Variability</i>		
Number of different tours		Increases with n (MacGregor & Ormerod, 1996)
H	$\sum_{(a,b) \in P^2} p_{ab} (-\log_2 p_{ab})$	Increases with n (MacGregor & Ormerod, 1996; Vickers et al., 2001)
H_N	$\frac{H}{H_{max}}$	Decreases with n (recomputed here using the data from MacGregor & Ormerod, 1996, and Vickers et al., 2001)
<i>Violations of Optimality Rules</i>		
Rule 1	Straight lines only	Violations typically due to limited motor precision; in rare cases associated with non-physical crossings (van Rooij et al., 2003)
Rule 2	No retraversals of edges	Violations often precluded by instructions (MacGregor et al., 1999; Vickers et al., 2001)
Rule 3	No revisiting of points	Violations often precluded by instructions (MacGregor et al., 1999; Vickers et al., 2001)
Rule 4	No crossing edges	Violations relatively rare, occurring in no more than about 6% of the cases (van Rooij et al., 2003)

A more commonly used measure of tour quality is the *normalized deviation from optimal*, defined as:

$$d_N = \frac{L_{obs} - L_{opt}}{L_{opt}} \quad (2)$$

where L_{opt} denotes the length of the optimal tour for the instance under consideration. It is consistently found that d_N scores systematically increase with instance size (Graham et al, 2000; MacGregor & Ormerod, 1996; MacGregor et al., 1999). Some authors have drawn particular attention to the fact that the increase in d_N seems to be roughly linear. For example, MacGregor et al. argued, "This is an important outcome, since the complexity of a TSP, in terms of the number of possible solutions [i.e., tours], increases as a *factorial* function of the number of points" (MacGregor et al., 1999, p. 1421, emphasis in original; see also p. 1425). We remark that this statement is misleading because there is a crucial difference between the number of possible tours and their average length. We performed an analysis and a Monte Carlo simulation study and found that the expected value of d_N for random tours (i.e., tours that visit points in random order) on randomly generated instances (i.e., where points are sampled uniformly within a bounded area) in fact increases more slowly rather than linearly with instance size. For details of this analysis, see [Appendix A](#).

The above-mentioned concern suggests that a more meaningful description of the instance size effect may be obtained when we evaluate the standardized tour length relative to the expected deviation from optimal tour length of a random tour:

$$z_d = \frac{L_{obs} - \mu_{rand}}{\sigma_{rand}} - \frac{L_{opt} - \mu_{rand}}{\sigma_{rand}} = \frac{L_{obs} - L_{opt}}{\sigma_{rand}} \quad (3)$$

which is equivalent to the *standardized deviation from optimal* used by Vickers et al. (2001). Vickers et al. observed that participants' mean z_d did *not* systematically increase with instance size. In fact, they observed a nonlinear trend but suggested the result might be spurious.

Finally, another commonly used measure is the proportion of instances for which a participant finds an *optimal* tour (or, alternatively, the proportion of participants who find an optimal tour for a given instance), denoted p_{opt} . It is consistently found that p_{opt} decreases relatively slowly with instance size (Graham et al., 2000; MacGregor et al., 1999, 2000; Vickers et al., 2001). Because the number of possible tours grows factorially, the relatively slow decrease in p_{opt} with instance size should perhaps not be regarded as a decline in performance at all.

The preceding illustrates that how quality of performance is affected by instance size seems to depend critically on how one defines quality. We show that a similar conclusion also holds for measures of between-subject variability.

Between-Subjects Variability

Based on the notion that “good patterns have few alternatives,” MacGregor and Ormerod (1996, p. 530) proposed that variability in the tours produced by different people for a given instance might index the difficulty of that instance for humans. Although difficulty is arguably better measured by the extent to which participants succeed (or fail) in producing tours of minimum length, we agree that between-subject variability is of interest as an index of the degree to which human tour construction is constrained by instance-specific properties. As a first possible measure of between-subject variability, MacGregor and Ormerod (1996) simply counted the number of different tours produced by participants for a given instance. They found that this number was on average larger for their 20-point instances than for their 10-point instances. Because this measure is insensitive to the degree of difference between tours, MacGregor and Ormerod also computed response uncertainty. Response uncertainty H is analogous to Shannon’s information theoretic formula (Shannon, 1948) for entropy, and is defined as:

$$H = \sum_{(a,b) \in P^2} p_{ab} (-\log_2 p_{ab}) \quad (4)$$

where P^2 denotes the set of pairs of points in the point set P , $a \neq b$, and p_{ab} denotes the probability that edge (a,b) is in a tour produced by a participant. Here p_{ab} is estimated by the proportion of tours produced by a group of participants that contain edge (a,b) .

MacGregor and Ormerod (1996) and Vickers et al. (2001) found that H tends to increase with instance size. The interpretation of this finding is complicated, however, by the fact that H increases with instance size even if the set of tours produced by the participants is a *random* sample from the set of all possible tours on that instance. Namely, assuming that the tour visits every point exactly once, we know that, for a random tour, $p_{ab} = \frac{n}{N}$ for all $(a,b) \in P^2$, where n denotes the number of points in P and $n = \binom{n}{2}$, and thus the response uncertainty over a set of random tours is given by:

$$H_{\max} = n \left(-\log_2 \frac{n}{N} \right) \quad (5)$$

Clearly, H_{\max} is also an increasing function of instance size (n).

We propose to use normalized response uncertainty, H_N , instead of H , when analyzing the effect of instance size:

$$H_N = \frac{H}{H_{\max}} \quad (6)$$

Note that $H_N = 1$ means that $H = H_{\max}$, which indicates a maximum dispersion of observed tours, whereas $H_N = 0$ means that $H = 0$, and thus, all observed tours are identical. Using Equations (5) and (6), we normalized the H -scores reported by MacGregor and Ormerod

(1996) and Vickers et al. (2001). We found that for the 10- and 20-point instances of MacGregor and Ormerod (1996) H_N scores were on average 0.26 and 0.22, respectively, and for the 10-, 25-, and 40-point instances of Vickers et al. (2001), the H_N scores were on average 0.24, 0.16, and 0.16, respectively. Interestingly, this pattern of results suggests that response uncertainty decreases, rather than increases, with instance size.

Violation of Optimality Rules

Although E-TSP is a computationally intractable task, some simple rules can be used to constrain the space of candidate solutions. First, because the shortest connection between any two points in the Euclidean plane is a straight line, an optimal tour will always consist of straight-line segments only (Rule 1). Further, because E-TSP requires each point (other than the starting point) to be visited no more than once, the number of straight-line segments in an optimal tour is exactly the same as the number of points in the set. In other words, an optimal tour never retraverses an edge (Rule 2) or revisits a point (Rule 3).³ Finally, an optimal tour also never contains line segments that cross in the plane (Rule 4; see also van Rooij et al., 2003). To understand these rules, no deep mathematical insights are required; one may even contend that they are obvious. Note, however, that the ability to see the validity of these rules does not imply an ability to apply this knowledge successfully to the task of E-TSP.

Research shows that no more than about 6% of the tours produced by adult participants violate one or more of Rules 2–4. Often, people do not draw perfectly straight lines (hence, strictly violating Rule 1), but this is probably due to limited precision in motor control. However, sometimes clearly *curved* line segments are observed in cases where straight-line segments would lead to violations of Rule 4. Such situations are called non-physical crossings and occur in fewer than 1% of all tours (see van Rooij et al., 2003, for an overview).

Given the large number of possible tours for any given instance that violate at least one of these rules, the low frequency of such violations means that humans solve the E-TSP in such a way that tours violating Rules 1–4 are precluded. This observation allows for at least two possible interpretations. First, it may mean that adult participants adopt a cognitive-analytic strategy that embodies Rules 1–4 in some form. Alternatively, it may be that tours that are perceived as good form happen to be those tours that do not violate these rules. MacGregor and colleagues seem to argue the latter position (e.g., MacGregor et al., 1999). If their position were correct then we would expect children to violate Rules 1–4 as infrequently as adults.

We briefly note that some researchers explicitly instruct their participants not to produce tours that violate Rules 2 and 3. For example, Vickers et al. (2001, p. 36; see also MacGregor et al., 1999) instructed their participants to produce tours that “passed through each point once and only once.” Note that this instruction is logically (though

not necessarily psychologically) redundant. In our experiment we did not explicitly instruct participants to produce tours that visit points exactly once. This allows us to test whether children and/or adults spontaneously comply with Rules 2 and 3.

Overview

We set out to investigate how performance on E-TSP varies with cognitive development (operationalized by different age groups) and to evaluate how instance size affects performance as measured by the different dependent measures discussed above. In doing so, we aim to distinguish between those aspects of performance that can be understood in terms of perceptual processing and those aspects that indicate cognitive-analytical processing. The use of many different dependent measures has several advantages. First, it allows us to test whether previously reported effects are replicated in our adult group. This not only serves to strengthen existing findings, but also serves to validate the use of our adult group as a control group. Secondly, it allows us to pick up on subtle, quantitative and qualitative differences between age groups. Finally, bringing all these measures together in one experiment also allows us to address the paradoxical effects of instance size discussed above.

The remainder of this article is organized as follows. We start in the following section on methods by describing relevant details of our empirical study: the participants, the stimulus material used in our experiment, and the experimental procedure. The next section reports on the results of the experiment. In turn, it discusses findings for three different classes of dependent measures: measures of tour quality, between-subject variability, and violations of optimality rules. We conclude with a general discussion in which we explain how our results support both the hypothesis that humans capitalize on natural perceptual organization of point sets in constructing solutions to the E-TSP and the hypothesis that high-quality performance on the E-TSP by adult participants reflects the use of cognitive-analytical skill.

Methods

Participants

A total of 59 participants volunteered for the experiment. Nineteen of the volunteers were grade 2 students ($M = 7.43$ years, $SD = 0.28$), referred to as the Age7 group. Twenty-three of the volunteers were grade 7 students ($M = 12.53$ years old, $SD = 0.29$), referred to as the Age12 group. Participants in these two groups attended the same elementary school and received a sticker for their participation. The remaining 18 volunteers ($M = 19.10$ years old, $SD = 0.65$) were students at the University of Victoria enrolled in an introductory psychology class who participated in the experiment for course credit. This last group will be referred to as the Adult group. The experiment was undertaken with

the understanding and consent of each participant and the understanding and consent of the parents of each child participant.

Stimuli

The stimuli were 15 randomly and independently generated point sets, 5 for each of three different instance sizes, $n = 5, 10, \text{ or } 15$ (refer to [Appendix B](#) for the x, y coordinates for the point sets used in this experiment). The x and y values of each point (x, y) were randomly and independently sampled from a uniform distribution on the interval $[0, 1]$, and scaled to an 8.7×8.7 cm square. All point sets were presented on paper (cut to $14 \text{ cm} \times 11 \text{ cm}$). The points were 2 mm in diameter. The only restriction placed upon the randomly generated instances was that no two points were to touch when presented on the paper (i.e., each point needed to be clearly visually distinguishable).

Procedure

All participants in a given age group were tested together in a classroom. The participants were first given illustrated instructions specifying that, for each instance, the participant's task was to draw a line that connected all the points such that (a) the line started and ended at the same point and (b) the total length of the line was as short as possible. Participants were given an analogy in the form of the following story:

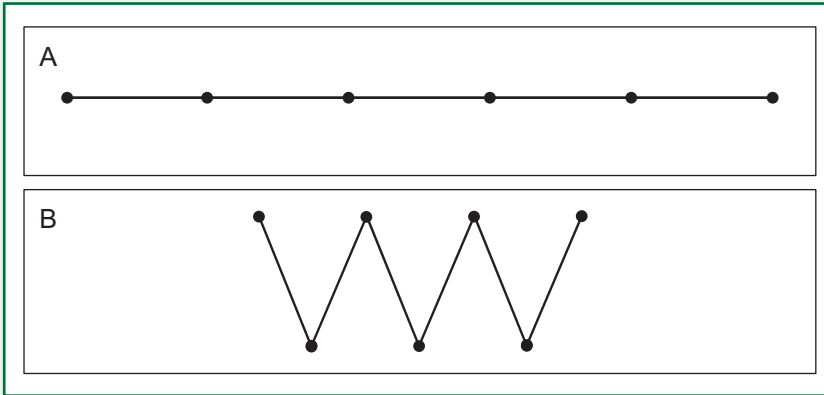
Each point represents a house, with one point representing your house. You have been invited to visit all of your friends, and each one lives in one of the houses. You want to go see each of your friends, but you do not want to have to do any extra walking because you will be tired. So you want to choose the path with the overall shortest distance. You start at your own house, and of course you must end up back at your own house, making a complete tour.

The instructions were identical for the three age groups, with the following exception. To help the Age7 group understand the concept of overall length, participants in this group were presented with Piaget's Matchstick Problem (see, e.g., Craig, Kermis, & Digdon, 2001), using straws instead of matchsticks (see [Figure 1](#) for an illustration). There were two sets of straws. The first was composed of five straws joined together in a straight line. The second was composed of six straws joined together at 45-degree angles making a two-dimensional zigzag line. The straws were held up in front of the children, one above the other. The row of straight straws started and ended farther apart than the row of bent straws (i.e., the beginning to the end of the straight row was farther apart than the beginning to end of the bent straws). The correct answer to the matchstick problem was discussed, and the children were given a full explanation. This procedure was used in

the Age7 group to ensure, as best as possible, that these participants understood what was meant by total length in the instructions for the task of E-TSP.

Figure 1.

Illustration of the Matchstick problem. The 7-year-old participants were shown that the total length of path A was shorter than the total length of path B.



Each participant was given a booklet containing the 15 instances (each presented on a separate page). To control for practice effects, the order of the instances in the booklets was randomized as follows. The 15 pages of the booklet were divided into five blocks, each block consisting of three consecutive pages. The instances were randomly assigned to a block with the constraint that each block contained one instance of each size (i.e., one 5-point instance, one 10-point instance, and one 15-point instance). Within each block, the instances were randomly ordered (i.e., 5, 10, 15 points, or 5, 15, 10 points, etc.) Within each age group, the order of instances differed for each participant. Across age groups, participants were randomly matched (as much as possible given the unequal sample sizes) so that the same order of presentation would be shown once to an Age7 participant, once to an Age12 participant, and once to an Adult participant.

Participants were instructed to solve the instances in the order that they appeared in their booklet. Participants used pencil (or pen) to draw the tours, and were allowed to use erasers (or receive a new copy of the instance) in order to revise their tours. Upon completion of the booklet, the experimenter checked each instance to ensure that every point was visited at least once and that a closed path (i.e., a tour) was drawn. If either of these instructions was not followed for a particular instance, the participant was asked if s/he wanted to redo the task for that instance. If the participant agreed, s/he was given a new copy of the respective point set. Testing was self-paced and occurred in one session (for each age group) lasting between 30 to 60 minutes.

Results

Data from one Age7 participant were excluded from analysis because the unsystematic nature of this participant's tours suggested that s/he was not following task instructions. Of the remaining 870 tours, seven (by six different participants) failed to visit every point in the point set. These were instances 8 and 13 in the Age7 group; instances 8, 10, and 12 in the Age12 group; and instances 2 and 3 in the Adult group. These tours were excluded from the analyses. For analyses requiring tour length estimates for every cell in the design (as in the ANOVAs reported below), we replaced the seven empty cells by the relevant age-group average tour length for that particular instance.

Measures of Tour Quality

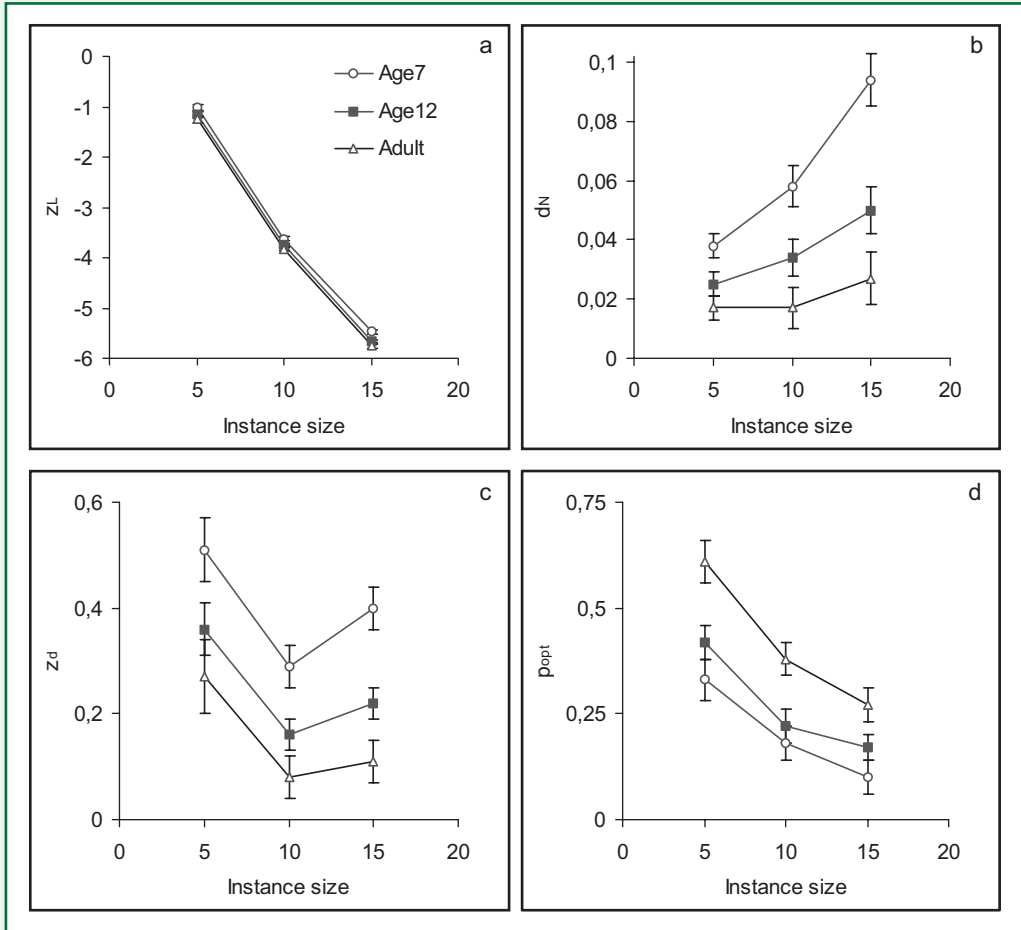
For each observed tour, tour length was computed in the unit square (with precision to four decimal places). Although participants did not always draw perfectly straight line segments to connect the points, tour length was always computed as the minimum distance required to visit the points in the order visited by the drawn tour. Some tours revisited points and retraversed edges in the tour (see also the section on [violations of optimality rules](#)). For those tours the order in which points were visited was not always unambiguously implied by the line drawing. In these ambiguous cases we gave the participant the benefit of the doubt and computed the shortest possible tour over the drawn line segments.

Because the experimental design was a $3 \times 3 \times 5$ mixed-factorial design, with one between-subjects variable AgeGroup (with three levels, Age7, Age12, and Adult) and two within-subjects variables of Size (three levels with $n = 5, 10,$ and 15) and Instance (5 instances for each Size), the appropriate statistical analysis to assess how each of these variables affected the measures z_L , d_N , and z_d , is an analysis of variance (ANOVA) for mixed designs (see, e.g., Kirk, 1995). However, because effects of Instance are not of interest, we do not report them (they are merely included in the analyses for statistical control). For the p_{opt} measure, which is the proportion of the five instances on which an optimal response was given, the appropriate analysis is a 3 (AgeGroup) \times 3 (Size) mixed-factorial ANOVA, with Size as the repeated measure. Where necessary, post-hoc analyses were performed on pairs of means, with an overall Type I error rate held at .05 and controlled by using the Bonferroni adjustment: This led to all post-hoc comparisons being tested with a per-comparison α adjusted to 0.025.

For each of the dependent measures of tour quality, z_L , d_N , z_d , and p_{opt} Figure 2 presents an overview of the results with the mean and standard error of the means plotted for each condition. We will discuss the results for each measure of tour quality in turn.

Figure 2.

Measures of tour quality plotted against instance size (5, 10, 15 points) for each of the age groups (Age7, Age12, Adult) separately. (a) Plots mean standardized tour length (z_L); (b) plots mean normalized deviation from optimal (d_N); (c) plots mean standardized deviation from optimal (z_d); (d) plots mean proportion of optimal tours (p_{opt}). Error bars represent the standard error of the mean (for the measure z_L , the error bars are too narrow to be clearly visible).



Standardized Tour Length

For each 5-point instance ($n = 5$), we computed the σ_{rand} and σ_{rand} of the 12 possible tours. Because the number of possible tours is very large for 10- and 15-point instances (recall that the number of tours on n points is given by $\frac{(n-1)!}{2}$), we estimated μ_{rand} and σ_{rand} for these instances based on random samples of 10,000 possible tours. Using these values in Equation 1, the standardized tour length, z_L , was computed for each observed tour. Figure 2a shows that, in terms of z_L , performance is relatively good for all age groups, with all age groups producing tours that are on average more than three standard deviations below the expected length of a random tour. The ANOVA on the z_L scores

revealed a significant main effect of AgeGroup, $F(2, 56) = 18.46, p < .001$, partial $\eta^2 = 0.397$, with z_L scores being significantly higher for the Age7 group ($M_{\text{Age7}} = -3.363, SEM = .029$) than for the Age12 group ($M_{\text{Age12}} = -3.518, SEM = .026$) and significantly higher for the Age12 group than for the Adult group ($M_{\text{Adult}} = -3.610, SEM = .029$). The assumption of sphericity was violated for the factor Size, but because the corrected statistics supported exactly the same conclusions as the uncorrected statistics, we simply report the uncorrected results here. The main effect of Size was found to be significant, $F(2, 112) = 9160.0, p < .001$, partial $\eta^2 = 0.994$, with z_L being significantly higher for the 5-point instances ($M_5 = -1.134, SEM = .031$) than for the 10-point instances ($M_{10} = -3.733, SEM = .021$), and the 10-point instances significantly higher than 15-point instances ($M_{15} = -5.625, SEM = .021$). There was no significant AgeGroup \times Size interaction, $F < 1$. In sum, performance in terms of z_L gets better with instance size and does so for all three age groups equally. Also, performance gets systematically better with age.

Normalized Deviation from Optimal

Optimal tour lengths were computed for each instance using a branch-and-bound algorithm (see, e.g., Papadimitriou & Steiglitz, 1998) implemented in Java (to obtain a copy of the program send an e-mail to stege@cs.uvic.ca). Then the normalized deviation from optimal, d_N , was computed using Equation 2 for each observed tour. The ANOVA on the d_N scores revealed a significant main effect of AgeGroup, $F(2, 56) = 22.73, p < .001$, partial $\eta^2 = 0.448$ (with $M_{\text{Age7}} = 0.063, SEM = .005$; $M_{\text{Age12}} = 0.036, SEM = .004$; and $M_{\text{Adult}} = 0.020, SEM = .005$), a significant main effect of Size, $F(2, 112) = 18.94, p < .001$, partial $\eta^2 = 0.253$ (with $M_5 = 0.026, SEM = .002$; $M_{10} = 0.036, SEM = .004$; and $M_{15} = 0.057, SEM = .005$), and a significant AgeGroup \times Size interaction, $F(4, 112) = 3.33, p < .05$, partial $\eta^2 = 0.106$. To probe the interaction we tested the effect of Size for each age group separately. The simple main effect of Size was present for each age group (though only marginally significant for the Adult group), but its effect size decreased systematically with age (partial $\eta^2 = 0.361, 0.244$, and 0.123 for the Age7, Age12, and Adult groups respectively).

Standardized Deviation from Optimal

Using Equation 3, the standardized deviation from optimal, z_d , was computed for each observed tour. The ANOVA on these z_d scores indicated a significant main effect of AgeGroup, $F(2, 56) = 18.46, p < .001$, partial $\eta^2 = 0.397$. Pairwise comparisons showed that z_d scores were significantly higher for the Age7 group ($M_{\text{Age7}} = .400, SEM = .029$) than for the Age12 group ($M_{\text{Age12}} = .245, SEM = .026$) and significantly higher for the Age12 group than for the Adult group ($M_{\text{Adult}} = .153, SEM = .029$). Again the assumption of sphericity

was violated for the factor Size, but as before, we report the uncorrected results. The main effect of Size was significant, $F(2, 112) = 18.43, p < .001$, partial $\eta^2 = 0.248$. Interestingly, the nature of the Size effect was somewhat counterintuitive: Although z_d scores were significantly higher for the 5-point instances ($M_5 = .376, SEM = .031$) than for the 15-point instances ($M_{15} = .244, SEM = .021$), the mean of the 10-point instances was significantly smaller ($M_{10} = .177, SEM = .021$) than that of both the 5- and 15-point instances. This general nonlinear relationship between z_d and instance size was replicated for all age groups (see also Table 1). There was no AgeGroup \times Size interaction, $F < 1$. Thus, consistent with both the analyses of z_L scores and d_N scores, the analysis of z_d scores shows that older participants performed better than younger participants. Inconsistent with those analyses, however, performance was best for the 10-point instances (or at least not different for 10- and 15-point instances). This nonlinear relationship between instance size and z_L scores mirrors Vickers et al.'s (2001) unexpected result and will be discussed later.

Proportion of Optimal Tours

For each instance size, we computed the proportion of instances (out of the five) for which a participant found the optimal tour. The ANOVA on these p_{opt} scores revealed a significant main effect of AgeGroup, $F(2, 56) = 16.94, p < .001$, partial $\eta^2 = 0.377$. The Age7 participants ($M_{Age7} = .204, SEM = .027$) found optimal tours less frequently than the Age12 participants ($M_{Age12} = .270, SEM = .024$), and the Age12 participants found optimal tours less frequently than the Adult participants ($M_{Adult} = .419, SEM = .027$). Further, the main effect of Size was significant, $F(2, 112) = 42.29, p < .001$, partial $\eta^2 = 0.430$. Not surprisingly, optimal tours were found significantly more often for 5-point instances ($M_5 = .454, SEM = .027$) than for 10-point instances ($M_{10} = .258, SEM = .022$), and more often for 10-point instances than for 15-point instances ($M_{15} = .180, SEM = .020$). Again, no significant interaction between AgeGroup and Size was found, $F < 1$.

Between-Subjects Variability

In these analyses, variability in participants' performance on the 15 instances is examined as a function of participants' age and instance size. For each instance, the following three measures were obtained: (a) P_{diff} , the number of different tours produced by different participants (divided by the number of participants contributing to the count), (b) H , the raw response uncertainty computed using Equation 4, and (c) H_N , the normalized response uncertainty (see Equations 5 and 6), which controls for the fact that response uncertainty increases with instance size even for random tours.

In Equation 4, to accommodate the fact that some people traversed some edges twice, we treated first-time and second-time traversals of an edge as two different events; that is, a second traversal of an edge (a,b) was coded as the traversal of a hypothetical edge (a',b') . For each possible first-time and second-time edge for a given instance the proportion of tours in a given age group containing that edge was computed. These proportions served as the probabilities to compute response uncertainty (zero probabilities were taken to add zero uncertainty). Treating first- and second-time traversals as the same kind of event would inflate response uncertainty. Namely, response uncertainty would, for example, be lower if half of the participants traversed an edge (a,b) once and the other half traversed it twice than it would be if all participants traversed the edge exactly once. Clearly, in the latter situation there is a stronger consensus among participants.

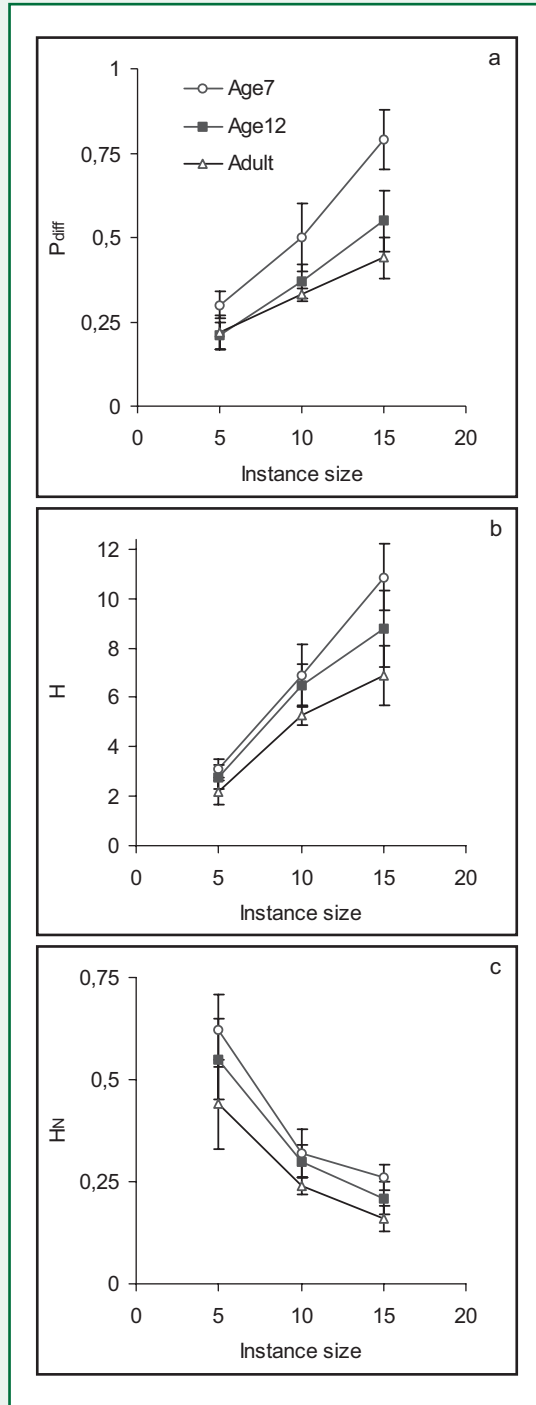
For H_N and Equation 6, despite the fact that participants sometimes visited points more than once, for simplicity we chose to base the normalization constant H_{max} on random tours that visit each point exactly once. The alternative of computing H_{max} based on random tours that can visit points more than once would make H_{max} grow even faster with n and, hence, would only serve to strengthen the observed effect of instance size for H_N .

For each of these measures, a 3 (AgeGroup) \times 3 (Size) mixed-factorial ANOVA was performed, with AgeGroup as the repeated measure on the instances and Size the between-instance factor. Figure 3 presents an overview of the results. All post-hoc comparisons were tested with a per-comparison α adjusted to 0.025.

For all three of these measures, there was a strong and significant main effect of AgeGroup, with the youngest participants producing more different tours and exhibiting the largest amount of response uncertainty, that decreased with increasing age. Specifically, for P_{diff} , $F(2, 24) = 18.11$, $p < .001$, partial $\eta^2 = 0.601$, with the Age7 group producing significantly more different tours ($M_{Age7} = 0.531$, $SEM = .069$) than the Age12 and Adult groups, which did not differ ($M_{Age12} = 0.377$, $SEM = .051$; $M_{Adult} = 0.333$, $SEM = .035$). For both response uncertainty measures, H and H_N , uncertainty decreased with age: For H , $F(2, 24) = 14.01$, $p < .001$, partial $\eta^2 = 0.539$, and for H_N , $F(2, 24) = 6.10$, $p < .01$, partial $\eta^2 = 0.337$.

Figure 3.

Measures of between-subject variability plotted against instance size (5, 10, 15 points) for each of the age groups (Age7, Age12, Adult) separately. (a) Plots the mean number of different tours divided by the number of participants (P_{diff}); (b) plots mean response uncertainty (H); (c) plots mean normalized response uncertainty (H_N). Error bars represent the standard error of the mean.



All three measures also showed a strong and significant main effect of instance size. Not surprisingly, for P_{diff} the number of different tours produced by the participants increased with instance size, $F(2, 12) = 10.03$, $p < .01$, partial $\eta^2 = 0.626$, although only the difference between M_5 and M_{15} reached significance. Given the fact that the number of possible tours grows factorially with instance size, the slow (almost linear) increase in P_{diff} is quite remarkable. Correspondingly, response uncertainty, H , also increased across instance size, $F(2, 12) = 11.70$, $p < .01$, partial $\eta^2 = 0.661$. For H_N , however, the significant main effect of Size, $F(2, 12) = 9.63$, $p < .01$, partial $\eta^2 = 0.616$, showed a different pattern: while H_N did not differ for 10- and 15-point instances ($M_{10} = .286$, $SEM = .024$, and $M_{15} = .210$, $SEM = .020$), it was significantly larger for the smallest, 5-point instances ($M_5 = .537$, $SEM = .056$).

The AgeGroup \times Size interaction was marginally significant for two of the three measures: P_{diff} [$F(4, 24) = 2.66$, $p < .10$, partial $\eta^2 = 0.307$] and H [$F(4, 24) = 2.72$, $p < .10$, partial $\eta^2 = 0.312$]. For both of these measures, the differences between age groups were somewhat larger for the larger instances, likely reflecting the fact that the number of possible solutions grows factorially with instance size. Notably, the AgeGroup \times Size interaction was not significant for the normalized measure, H_N , $F < 1$.

Violations of Optimality Rules

Each of the observed 863 tours was coded for the frequency of occurrence of four kinds of violations of the optimality rules: (a) two drawn line-segments crossed in the plane at a point that is not part of the instance; (b) two line-segments would have crossed in the plane had the participant drawn them as straight line-segments (called a non-physical crossing); (c) a point in the instance is revisited in the tour; and (d) an edge is retraversed in the tour. Because (a)–(d) cover all situations in which a tour intersects itself in one or more points, we call tours for which none of (a)–(d) applies *crossing-free*. Recall that an optimal tour is always crossing-free, and thus (a)–(d) are all violations of this optimality rule. Figure 4 illustrates each type of crossings and Table 2 summarizes the frequency with which they occurred.

Figure 4.

Illustrations of the different types of crossings.

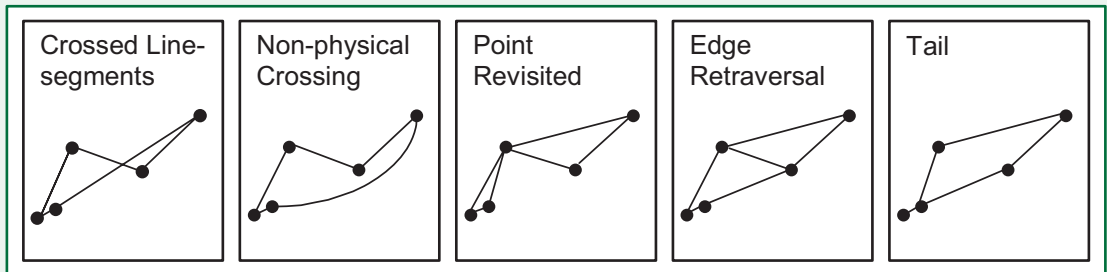


Table 2

For Each Age Group (Age7, Age12 and Adult) and for Each Instance Size ($n = 5, 10, \text{ or } 15$), the Percentage of Tours that are Crossing-free and the Frequency with which a Particular Type of Crossing Occurred (and in How Many Tours).

				Crossed line-segments	Non-physical crossings	Revisited points	Retraversed edges
AgeGroup	Size	Total no. of tours	Crossing-free tours (%)	Freq. (tours)	Freq. (tours)	Freq. (tours)	Freq. (tours)
Age 7	5	90	82.2	5 (5)	8 (6)	8 (7)	7 (7)
	10	89	83.1	7 (6)	5 (3)	12 (6)	8 (4)
	15	89	76.4	4 (2)	10 (9)	28 (11)	16 (9)
Age12	5	115	87.0	3 (3)	3 (3)	9 (9)	8 (8)
	10	113	93.8	0 (0)	1 (1)	12 (6)	11 (6)
	15	114	96.5	1 (1)	0 (0)	10 (3)	10 (3)
Adult	5	88	95.5	1 (1)	0 (0)	3 (3)	3 (3)
	10	90	95.6	1 (1)	0 (0)	3 (3)	1 (1)
	15	90	97.8	0 (0)	0 (0)	3 (2)	3 (2)

All age groups produced many more crossing-free tours than could be expected by chance alone, indicating that participants in all age groups tend to avoid crossings. However, there were age differences: The Age7 group produced substantially more tours with crossings (19.4%) than the Age12 group (7.6%), and the Age12 group in turn more than the Adult group (3.7%). A significant percentage of crossings were revisited points: 55% (Age7), 80% (Age12), and 82% (Adult). The majority of these revisited points also involved edge retraversals; 64% (Age7), 94% (Age12), and 77% (Adult). Interestingly, most retraversals were part of what we call a *tail*: If points p_1, p_2, \dots, p_k , with $k \leq 1$, are visited twice by first traversing the edges $(p_1, p_2), (p_2, p_3), \dots, (p_{k-1}, p_k)$ in order and then immediately (without visiting any other points first) retraversing the same edges in the reversed order, we call the sequence of edges $(p_1, p_2), (p_2, p_3), \dots, (p_{k-1}, p_k)$ a tail (see Figure 4 for an example). As many as 55% (Age7), 97% (Age12), and 100% (Adult) of the edge retraversals were part of a tail. Finally, we also note that the young participants produced non-physical crossings (29% of all crossings for Age7 group and 10% for the Age12 group), whereas Adults did not (0%).

We remark that our experiment has, as a side-product, produced some results that are relevant for the continuing debate about the role of the convex hull in human performance on E-TSP (see, e.g., MacGregor & Ormerod, 1996, 2000; Lee & Vickers, 2000; van Rooij et al., 2003). This rather specialized debate falls outside the main scope of this

paper, and the interested reader is referred to Appendix C for the presentation and discussion of those results.

Conclusion

With our experiment we set out to address two general questions. Is overall level of performance on E-TSP a function of level of cognitive-analytic development, and how does quality of performance at each level of cognitive-analytic development scale with instance size? We observed that even the youngest children produced tours that were, on average, more than three standard deviations shorter than a random tour and within about 7% of the optimal tour length. Further, even for the 15-point instances, the youngest children found the optimal tour as often as 10% of the time, on average, compared to a $1 \text{ in } 4 \times 10^{10}$ chance of finding it by chance. The performance of young children shows that high quality of performance is attainable with little or no analytical processing. This finding supports MacGregor, Ormerod, and Chronicle's hypothesis that the task of E-TSP coincides with natural perceptual tendencies in humans (MacGregor & Ormerod, 1996, MacGregor, Ormerod, & Chronicle, 1999). On the other hand, we have also found significant and systematic improvements in performance with age, with older participants producing on average shorter tours than younger participants. This finding suggests that older participants are successfully exploiting their analytical resources to perform the task better than younger participants.

We investigated how often participants produced tours containing violations of different types of optimality rules, collectively called "crossings". We found that even the youngest children produced relatively few tours with crossings (about 19.4% of all tours). This finding suggests that a high level of crossing-avoidance is attainable with little or no analytical processing and is consistent with the idea that perceived good forms tend be crossing-free. At the same time, however, we found that the number of tours with crossings systematically decreased with age, indicating that older participants use strategies that go beyond perceptual processing to avoid crossings in tours.

We have replicated the effects of instance size for all dependent measures reported in the literature in the adult group. Interestingly, all these effects are also present in the data for the 7- and 12-year-old participants. Our finding that performance scales with instance size in a similar way across the age groups suggests that most cognitive processing on E-TSP is common across the age groups. Although we did find that the deviation from optimal (d_N) grew somewhat faster with instance size for the younger age groups and some qualitative differences in performance (types of crossings), we also found that, relative to the length and number of all possible tours for an instance, performance improved with instance size (viz., z_L increases; z_d does not increase; and p_{opt} grows relatively slowly with n) and did so for all groups equally.

We interpret the pattern of results as follows: Participants of all ages are able to exploit a computationally cheap and cognitively low-level process to significantly reduce the psychological problem space for E-TSP, hence avoiding much of the computational complexity inherent in the task. This explains how even 7-year-olds are able to attain a relatively high level of performance. Importantly, the nature of this process is such that the relative gain in problem-space reduction grows with instance size (also reflected in the decrease for all age groups in the normalized response uncertainty, H_N , as a function of instance size). Note that the perceptual process of finding good form, as hypothesized by MacGregor and colleagues, can perform the role of this hypothesized computationally cheap process because, arguably, the proportion of possible tours that are good form decreases drastically with instance size. Consider, as an illustration, the restriction that the good forms are (typically) crossing-free (see also Ormerod & Chronicle, 1999): For 5-point instances, about 28% of all possible tours are crossing-free, while for 10-point instances, no more than 0.2% of all possible tours are crossing-free. Given that there are likely more rather than fewer restrictions (i.e., not all crossing-free tours are necessarily good form), the proportion of tours that are good form decreases probably even faster with instance size.

Our interpretation also naturally reconciles the seemingly conflicting results of instance size reported in the literature and replicated here. On the one hand, the observation that performance (as evaluated relative to the length of optimal tours) worsens with instance size reflects that the cost of analytical processing increases simply because the complexity of the task grows (i.e., the size of the problem space grows with instance size). On the other hand, the observation that performance improves and response-uncertainty decreases (as evaluated relative to random tours) with instance size can be understood as an increasing payoff of perceptual processing with increasing n (i.e., the reduction in the problem space based on good form also increases with instance size).

In sum, we conclude that both MacGregor and colleagues and Vickers and colleagues are probably right: Perceptual processing, aided by little or no analytical skill, is sufficient to produce relatively high-quality solutions for visually presented E-TSP, as is evidenced by the high level of performance in children, but at the same time, adults seem to be using their analytical skills intelligently so as to improve upon what their visual systems provide. It seems to us that this kind of strategy may be adaptive not only for visual problem-solving tasks but also for complex optimization tasks in general. Possibly, the ability of humans to behave near-optimally in many complex real-world tasks may be similarly understood in terms of them intelligently capitalizing on computationally cheap, low-level processes provided to them by evolution and/or development (cf. Todd & Gigerenzer, 2000).

References

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Craig, G. J., Kermis, M. D., & Digdon, N. L. (2001). *Children today*. Toronto: Prentice Hall.
- Diaz, J., Petit, J., & Serna, M. (1998). Random geometric problems on $[0, 1]^2$. In M. Luby, J. Rolim, & M. Serna (Eds.), *Randomization and approximation techniques in computer science, number 1518 in Lecture Notes in Computer Science*. Heidelberg: Springer-Verlag.
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, 51(4), 380–417.
- Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness*. New York: W. H. Freeman and Company.
- Ginsberg, H. & Opper, S. (1979). *Piaget's theory of intellectual development*. Englewood Cliffs, NJ: Prentice-Hall.
- Graham, S. M., Joshi, A., & Pizlo, Z. (2000). The Traveling Salesman Problem: A hierarchical model. *Memory & Cognition*, 28(7), 1191–1204.
- Haselager, W. F. G. (1997). *Cognitive science and folk psychology: The right frame of mind*. London: Sage.
- Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Johnson, S. P. & Bremner, J. G. (2000). The role of good form in young infants' perception of partly occluded objects. *Journal of Experimental Child Psychology*, 76(1), 1–25.
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt Brace.
- Lazoff, D. M., & Sherman A. T. (1994). *An exact formula for the expected wire length between two randomly chosen terminals*. Technical Report TR CS-94-08, Computer Science Department, University of Maryland, Baltimore County.
- 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(1), 226–228.
- MacGregor, J. N., & Ormerod, T. (1996). Human performance on the Traveling Salesman Problem. *Perception & Psychophysics*, 58(4), 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(7), 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(7), 1183–1190.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman and Company.
- Nairne, J. S., Smith, M. S., Lindsay, D. S. (2001). *Psychology: The adaptive mind* (1st Canadian ed.). Toronto: Nelson Thompson.

- Oaksford M. & Chater, N. (1998). *Rationality in an uncertain world: Essays on cognitive science of human reasoning*. East Sussex, UK: Psychology Press Ltd., Publishers.
- Ormerod, T. C., & Chronicle, E. P. (1999). Global perceptual processing in problem solving: The case of the traveling salesperson. *Perception & Psychophysics*, 61(6), 1227–1238.
- Papadimitriou, C. H. & Steiglitz, K. (1998) *Combinatorial optimization*. New York: Dover Publications.
- Polivanova, N. I. (1974). On some functional and structural features of the visual-intuitive components of a problem-solving process. *Voprosy Psychologii [Questions in Psychology]*, 4, 41–51.
- Pylyshyn, Z., (Ed.) (1987). *The robot's dilemma: The frame problem in artificial intelligence*. Norwood: Ablex Publishing.
- Quinn, P. C. & Bhatt, R. S. (2005). Learning perceptual organization in infancy. *Psychological Science*, 16, 511–515.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.
- Simon, H. A. (1957). *Models of man: Social and rational*. New York: John Wiley & Sons.
- Simon, H. A. (1988). Rationality as process and as product of thought. In D. E. Bell, H. Raiffa, & A. Tversky (Eds.), *Decision making: Descriptive, normative, and prescriptive interactions*. (pp. 58–77), Cambridge: Cambridge University Press.
- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, 41(1), 1–19.
- Slater, A. (2000). Visual perception in the young infant: Early organization and rapid learning. In D. Muir & A. Slater (Eds.), *Infant development: The essential reading* (pp. 95–116). Malden, MA: Blackwell.
- Todd, P. M. & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23, 727–780.
- van Rooij, I., Stege, U., & Schactman, A. (2003). The convex hull and crossings in the Euclidean Traveling Salesperson Problem: Implications for human performance studies. *Memory & Cognition*, 31(2), 215–220.
- Vickers, D., Butavicius, M., Lee, M., & Medvedev, A. (2001). Human performance on visually presented Traveling Salesman Problems. *Psychological Research*, 65, 34–45.
- von Neuman, J. & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Author Note

The first and second authors contributed to this research equally. A substantial part of this paper is based on an unpublished thesis submitted by A. Schactman for the degree of B.Sc. (Honours Psychology) at the University of Victoria, Victoria, Canada. Parts of this research were presented at the Fourth Annual Meeting of the Northwest Cognition and Memory Society.

Acknowledgments

This research is supported by a research grant from the National Science and Engineering Research Council (NSERC) of Canada and a research grant from the University of Victoria awarded to U. Stege and by NSERC Grant OGP0121280-97 awarded to H. Kadlec. We are grateful to Michael Chow and Denise Broekman for their help with the data collection and analysis. We thank James Dixon and an anonymous reviewer for their helpful comments on an earlier version of this article. We also thank the administration, staff, parents, and children at Fairburn Elementary School for their contribution to this research.

Notes

1. A computationally intractable problem is a problem for which no algorithm exists that solves the problem in polynomial time (see, e.g., Garey & Johnson, 1979).
2. The *size* of an E-TSP instance is the number of points in the instance. This manipulation is often referred to as *problem size* in the psychological literature on E-TSP. To avoid confusion between “problem” and “instance,” we choose to use the term *instance size* instead. Following convention in computer science, we use the word *instance* to refer to a particular instantiation of the E-TSP problem (i.e., a particular set of points for which one is to find the shortest tour), and the word *problem* to refer to the generic problem, called E-TSP.
3. An edge retraversal implies that a point is revisited, but the reverse is not necessarily true.

Appendix A

Expectations for Random and Optimal Tours on Random Instances

First, we define the terms *random tour*, *random point*, and *random instance*. Let P be a point set consisting of n points, $P = \{p_1, p_2, \dots, p_n\}$. A random tour on P is a tour, T , that visits each point in P exactly once and in a random order (sampled without replacement). The last point visited is joined with the first point visited to complete tour T . A random point (in the unit square) is a point $p = (p_x, p_y)$, where the coordinates p_x and p_y are independent uniform random variables, with $0 \leq p_x, p_y \leq 1$. A random instance is a point set P where every point $p \in P$ is a random point.

We note that, if the area in which points can be located is fixed (as is the case in the psychological studies), then the expected tour length is a linearly growing function of n for random tours on random instances. Specifically, if T is a random tour and P is a random instance with $|P| = n$, then the expected length of T , $E(L_T)$, is a linearly increasing function of n . The basic rationale is that any tour with n points has n edges, with each

edge (p,q) having the same expected length, $E(\text{dist}_{pq})$. Because $E(L_T)$ is the sum of the expected values of the individual lengths of each edge, we get $E(L_T) = nE(\text{dist}_{pq})$. For a formal proof, see Lazoff and Sherman (1994). Furthermore, Lazoff and Sherman (1994) have shown that if the points in P are restricted to lie in the unit square $[0,1]^2$ then $E(\text{dist}_{pq}) \approx 0.5214$, and thus, $E(L_T)$ is approximately $0.5214n$.

For any given instance P , exactly one optimal tour length $(L_{opt} | P)$ exists and remains constant across all random tours on P . This means that the tour length of any random tour, T , on given random instance, P , is a random variable $(L_{obs} | P)$ that it is *independent* of $(L_{opt} | P)$. This observation is useful in deriving the expected value of the tour-length measure d_N for random tours T on random instances P , $E(d_N)$. Note that while $(L_{opt} | P)$ is constant for a given P , for all random instances of the same size n , $(L_{opt} | n)$ is itself a random variable across the P s.

From Equation 2 we have $d_N = \frac{L_{obs}}{L_{opt}} - 1$ for an observed tour of length L_{obs} . We next clarify the notation for L_{obs} : For a given random instance P and a given random tour T for P , $L_{obs} = L_T$. Further, we can write $(d_N | P, T) = \frac{(L_T | P)}{(L_{opt} | P)} - 1$. Next, we first determine $E(d_N | P, T)$, and then we derive $E(d_N)$ for all possible random tours T on all possible random instances P of a given size n on the unit square.

For a given random instance P and a given random tour T for P ,

$$E(d_N | P, T) = E\left(\frac{(L_T | P)}{(L_{opt} | P)} - 1\right) = \frac{E(L_T | P)}{(L_{opt} | P)} - 1 \quad (A1)$$

We could, if we wished, further reduce Equation A1 by noting that $E(L_T | P) = Cn$, where C is given above as approximately 0.5214 for the unit square. To take the expectation over all possible random instances of size n , we have:

$$E(d_N | n) = E_P\left(\frac{(L_T | P, n)}{(L_{opt} | P, n)} - 1\right) = \frac{E_P(L_T | n)}{E_P(L_{opt} | n)} - 1 \quad (A2)$$

because $(L_T | P)$ and $(L_{opt} | P)$ are independent.

We now consider how $E(L_{opt} | n)$ grows as a function of n . It has been shown that the expected tour length of optimal tours on a random instance of size n , $E(L_{opt} | n)$, converges in probability to some constant times the square root of n (Diaz, Petit, & Serna, 1998). That is, there exists a constant β , such that:

$$E(L_{opt} | n) \rightarrow \beta\sqrt{n}, \text{ for } n \rightarrow \infty, \text{ a.s. (almost surely)} \quad (A3)$$

Using this result in Equation A2, we conclude that,

$$E(d_N | n) = \frac{E(L_T | n)}{E(L_{opt} | n)} - 1 = \frac{cn}{\beta\sqrt{n}} - 1 = \frac{c}{\beta}\sqrt{n} - 1, \text{ as } n \rightarrow \infty, \text{ a.s.} \quad (\text{A4})$$

Thus, in the limit of n , $E(d_N)$ grows slower than a linear function of n , viz., as a square-root function of n .

Unfortunately, the limit in Equation A4 shows convergence in probability as n goes to infinity, but it does not exactly specify $E(L_{opt} | n)$ for finite n (Diaz, Petit, & Serna, 1998). Instance size in psychological experimentation is typically small, on the order of $5 \leq n \leq 40$. To get an idea for how $E(L_{opt} | n)$ grows as a function of n for small n , we performed a Monte Carlo simulation study to estimate the $E(L_{opt} | n)$. Because E-TSP is an NP-hard problem, we chose to limit the simulation to $n \leq 15$. For every n , a set of random instances on $[0,1]^2$ were generated; 10,000 instances for each $n \leq 10$, 1,000 instances for $n = 11, 12$, and 100 instances for $n = 13, 14, 15$. For each instance the optimal tour length was found using a branch-and-bound algorithm implemented in Java (to obtain a copy of the program send an e-mail to stege@cs.uvic.ca). Then, $E(L_{opt} | n)$ was estimated by computing the average optimal tour length over the 10,000 (or 1,000 or 100) random instances for each level of n .

The simulation results, given in Table A1, show that for these small, finite values of n , $E(L_{opt} | n)$ grows, at worst, linearly as a function of n . Note that there is a slight hint of a negative acceleration in these estimated functions, which is consistent with our finding in Equation A4 that as n goes to infinity $E(d_N)$ will increase with n , almost surely, as a square-root function of n .

Table A1

Expected Tour Length, $E(L_T | n)$, Expected Optimal Tour Length, $E(L_{opt} | n)$, and Expected Normalized Deviation from Optimal, $E(d_n | n)$, for Random Tours as a Function Instance Size, n , for Random Instances in the Unit Square. (Here $E(L_{opt} | n)$ and $E(d_n | n)$ are estimated by Monte Carlo simulation.)

N	5	6	7	8	9	10	11	12	13	14	15
$E(L_T n)$	2.61	3.13	3.65	4.17	4.69	5.21	5.74	6.26	6.78	7.30	7.82
$E(L_{opt} n)$	2.13	2.32	2.48	2.61	2.74	2.88	2.99	3.10	3.17	3.30	3.35
$E(d_n n)$	0.23	0.35	0.47	0.60	0.71	0.81	0.92	1.02	1.13	1.21	1.33

Appendix B

x,y -Coordinates in $[0,1]^2$ of the Points Sets in the Experiment.

Instance 1		Instance 2		Instance 3		Instance 4		Instance 5	
x	y	x	y	x	y	x	y	x	y
0.982	0.377	0.591	0.666	0.637	0.209	0.683	0.018	0.092	0.249
0.682	0.385	0.578	0.798	0.746	0.970	0.533	0.385	0.138	0.625
0.037	0.111	0.201	0.840	0.531	0.542	0.086	0.624	0.329	0.480
0.841	0.474	0.700	0.831	0.429	0.686	0.308	0.473	0.042	0.195
0.615	0.288	0.518	0.752	0.190	0.247	0.728	0.932	0.487	0.824

Instance 6		Instance 7		Instance 8		Instance 9		Instance 10	
x	y	x	y	x	y	x	y	x	y
0.439	0.322	0.453	0.232	0.451	0.108	0.850	0.583	0.024	0.707
0.831	0.439	0.419	0.349	0.166	0.686	0.988	0.661	0.746	0.337
0.554	0.563	0.858	0.676	0.822	0.093	0.761	0.443	0.985	0.636
0.452	0.110	0.445	0.036	0.079	0.360	0.768	0.133	0.767	0.389
0.599	0.569	0.101	0.120	0.736	0.646	0.382	0.581	0.597	0.861
0.251	0.380	0.107	0.252	0.845	0.747	0.454	0.021	0.976	0.158
0.034	0.584	0.238	0.217	0.454	0.978	0.235	0.112	0.883	0.184
0.357	0.095	0.750	0.832	0.327	0.791	0.921	0.040	0.144	0.011
0.307	0.553	0.634	0.828	0.159	0.046	0.796	0.131	0.462	0.089
0.550	0.505	0.741	0.588	0.004	0.920	0.290	0.136	0.501	0.076

Instance 11		Instance 12		Instance 13		Instance 14		Instance 15	
x	y	x	y	x	y	x	y	x	y
0.585	0.791	0.447	0.423	0.100	0.349	0.087	0.035	0.975	0.611
0.916	0.537	0.939	0.889	0.665	0.486	0.867	0.756	0.766	0.345
0.101	0.962	0.413	0.528	0.257	0.187	0.638	0.394	0.885	0.294
0.540	0.467	0.245	0.940	0.707	0.323	0.302	0.787	0.560	0.900
0.945	0.460	0.334	0.313	0.028	0.790	0.169	0.162	0.834	0.703
0.510	0.804	0.556	0.729	0.701	0.275	0.372	0.300	0.968	0.301
0.969	0.427	0.185	0.014	0.590	0.165	0.987	0.182	0.321	0.362
0.063	0.969	0.455	0.284	0.837	0.199	0.028	0.354	0.677	0.503

Instance 11		Instance 12		Instance 13		Instance 14		Instance 15	
x	y	x	y	x	y	x	y	x	y
0.519	0.994	0.287	0.652	0.698	0.768	0.556	0.141	0.350	0.787
0.689	0.625	0.624	0.977	0.466	0.267	0.606	0.944	0.884	0.195
0.932	0.300	0.444	0.986	0.333	0.166	0.679	0.155	0.781	0.759
0.341	0.611	0.203	0.112	0.022	0.834	0.013	0.780	0.318	0.749
0.007	0.769	0.807	0.487	0.436	0.856	0.012	0.516	0.709	0.817
0.174	0.227	0.758	0.214	0.858	0.669	0.187	0.137	0.868	0.437
0.222	0.043	0.527	0.699	0.635	0.908	0.950	0.964	0.753	0.389

Appendix C

The Role of the Convex Hull in Children's Performance on E-TSP

The *convex hull* of a point set in the Euclidean plane is the smallest convex polygon that contains all the points in the set. A point on the convex hull is called a *convex-hull point*. MacGregor and Ormerod (1996) have proposed that, because people can spot the convex hull via some automatic perceptual process and optimal tours always visit points on the convex hull in order of adjacency, people may use a strategy that specifically aims to produce tours that connect adjacent convex-hull points in order of adjacency (Prediction 1), as well as connect adjacent convex-hull points to each other (Prediction 2). Van Rooij et al. (2003) have argued that these hypotheses do not find support in the available data for adult participants. Given the presumed perceptual bases of these predictions, we would expect these predictions to be borne out at least in children's performance data. We report our relevant results below.

Test of Prediction 1

Because crossing-free tours necessarily visit points on the convex hull in order of adjacency, they do not bear on the hypothesis that people aim at producing tours that visit points on the convex hull in order of adjacency (see van Rooij et al., 2003, for a complete argument). Thus we need to specifically inspect tours with crossings to test this hypothesis. In our data, 52 tours for the Age7 group, 26 tours for the Age12 group, and 10 tours for the Adult group contained one or more crossings. Of these tours 19%, 58%, and 70% followed the convex hull for the Age7, Age12, and Adult groups respectively. We realize that these percentages are not very reliable for the Age12 and Adult groups, given the small numbers of cases. However, in the case of the Age7 group, clearly the majority of tours with crossings do not visit points on the convex hull in order of adjacency. If anything, this finding is evidence against the convex-hull hypothesis for the youngest age

group. Because performance in this group is presumably driven most by perceptual processing, this observation casts doubt on the idea that perceptual processing in E-TSP specifically induces tours that visit points on the convex hull in order of adjacency.

Test of Prediction 2

We first determined for each age group separately the average number of indentationsⁱ observed for each instance (I_{obs}). For each instance we computed the expected mean (I_{exp}) and standard deviation (σ_I) of the distribution of indentations for random tours, as well as the number of indentations in the optimal tour for that instance (I_{opt}) (Because instances 2 and 5 had only one interior point, and thus $\sigma_I = 0$, these instances were excluded from this analysis). Then we computed two indentation scoresⁱⁱ: (a) the standardized difference between observed and expected number of indentations, $D = (I_{obs} - I_{exp})/\sigma_I$ and (b) the standardized difference between observed and optimal number of indentation, $D_{opt} = (I_{obs} - I_{opt})/\sigma_I$.

We performed one-sample *t*-tests for each age group separately and found that D was significantly smaller than zero for the Age7 group ($M_{Age7} = -.487$, $SEM = .194$), $t(12) = -2.66$, $p < .05$, but not for the Age12 group ($M_{Age12} = -.351$, $SEM = .224$) or for the Adult group ($M_{Adult} = -.299$, $SEM = .230$), both $ps > .10$. Similarly, D_{opt} was significantly smaller than zero for the Age7 group ($M_{Age7} = -.515$, $SEM = .331$), $t(12) = -2.25$, $p < .05$. For the Age12 group and the Adult group, means of D_{opt} were marginally smaller than zero ($M_{Age12} = -.397$, $SEM = .176$; and $M_{Adult} = -.327$, $SEM = .128$), both $ps < .10$. Hence, despite the strictures of van Rooij et al. (2003), we find that our participants showed some tendency to connect adjacent convex-hull points to each other. In line with its presumed perceptual basis, this effect was most pronounced in the youngest age group.

Notes to Appendix C

- i. An *indentation* in a tour occurs every time the tour visits at least one interior point between two convex hull points.
- ii. Note that the finding by MacGregor and Ormerod (1996, Experiment 1) that participants produced tours with fewer indentations than expected by chance (*viz.*, on average, 1.8 *versus* 2.2 indentations, respectively) is confounded by the fact that their participants tended to find (close to) optimal tours (*viz.*, optimal tours for their instances had, on average, 1.5 indentations). To control for such a possible confound we compare the number of observed indentations to the number of indentations in both random and optimal tours.