

Perceived optimality of competing solutions to the Euclidean travelling salesperson problem

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Kyritsis, M. ORCID: <https://orcid.org/0000-0002-7151-1698>,
Gulliver, S. R. ORCID: <https://orcid.org/0000-0002-4503-5448>,
Feredoes, E. and Stouraitis, V. (2022) Perceived optimality of competing solutions to the Euclidean travelling salesperson problem. *Cognitive Systems Research*, 74. pp. 1-17. ISSN 1389-0417 doi: [10.1016/j.cogsys.2022.02.001](https://doi.org/10.1016/j.cogsys.2022.02.001) Available at <https://centaur.reading.ac.uk/103818/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.cogsys.2022.02.001>

Publisher: Elsevier

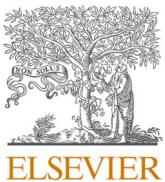
All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



Perceived optimality of competing solutions to the Euclidean travelling salesperson problem



Markos Kyritsis ^{a,*}, Stephen R. Gulliver ^a, Eva Feredoes ^b, Vasilios Stouraitis ^c

^a Henley Business School, Business Informatics Systems and Accounting, Informatics Research Centre, University of Reading, RG6 6UD, United Kingdom

^b School of Psychology and Clinical Language Sciences, University of Reading, Whiteknights Road, RG6 6AL, United Kingdom

^c Glasgow School for Business and Society, Glasgow Caledonian University, Cowcaddens Road, G4 0BA, United Kingdom

ARTICLE INFO

Action editor: Mog Stapleton

Keywords:

Travelling salesperson problem
Choice discrimination task
Monte Carlo methods
Perceptual activation
Geometric properties

ABSTRACT

The Travelling Salesperson Problem (TSP) is a nondeterministic-polynomial hard (NP-hard) combinatorial problem that occurs in a wide range of industrial domains, including logistics, route finding, and computer wiring. Interestingly, despite the problem's inherent computational difficulty, when presented in Euclidean space (ETSP), human participants can produce close-to-optimal solutions in near-linear time. However, when asked to compare and select the most optimum solution from a set of pre-defined competing solution options, participants can struggle. In this study we investigate this paradox by asking participants to compare four closed-loop Euclidean TSP solutions, in order to determine which solution they perceived to have the most optimal tour cost. We hypothesise that the extracted geometric properties have an effect on stimulus selection in a discrimination task (selection or no selection). Accordingly, we extracted four geometric properties from competing stimuli in order to create a perceptual activation function. Predictive analytics demonstrated that a classification model could identify the most optimal solution 97% of the time using the perceptual activation scores alone, yet human participants only correctly determined the most optimal solution 47% of the time. Mixed-effects models suggest that 'likelihood of stimulus selection' can be modelled as a function of the weighted coefficients of competing perceptual activation scores within each trial; however only a small amount of the variance is explained by these perceptual activation scores. Finally, a drift-diffusion model was used to create a theoretical framework of how likelihood of stimulus selection is influenced by competing perceptual activators. Our study highlights a novel way of extracting and analysing the importance of geometric properties that influence ETSP discrimination tasks, and links this analysis to human behaviour when discriminating between competing ETSP solutions.

1. Introduction

A travelling salesperson might plan to visit a list of cities before returning home. Naturally, the salesperson aims to (a) minimize the distance travelled (i.e., the tour cost), and (b) avoid visiting the same city twice. This classic problem, referred to as the Travelling Salesperson Problem (TSP), might seem simple, however solving this problem is deceptively complex due to the exponentially increasing number of possible routes – defined by the equation $(n-1)!/2$, where n is the number of cities. For example, a list of ten cities has 181,440 possible route solutions, yet adding just one more causes a ten-fold increase in possible routes to 1,814,400. For this type of combinatorial problem – a nondeterministic polynomial (NP) hard problem – exhaustive search

algorithms soon fail to produce optimal solutions within a viable time frame. Accordingly, applied algorithmic approaches focus instead on producing tours that are considered to be 'good enough', or 'near-optimal', with a practical trade-off between processing time (efficiency) and tour length (effectiveness) (Kyritsis, Gulliver, Feredoes, & Din, 2018b).

Interestingly, when presented in Euclidean space (ETSP) on a 2D graph, e.g., with dots/nodes representing cities, humans producing surprisingly highly efficient near-optimal tours in near-linear time (Dry, Lee, Vickers, & Hughes, 2006; Graham, Joshi, & Pizlo, 2000; Kyritsis et al., 2018b; MacGregor & Ormerod, 1996); a finding that also applies to navigational-variants of the TSP, i.e., in real space rather than on Euclidean 2D graphs (Blaser & Wilber, 2013). The lack of any significant

* Corresponding author.

E-mail address: m.kyritsis@henley.ac.uk (M. Kyritsis).

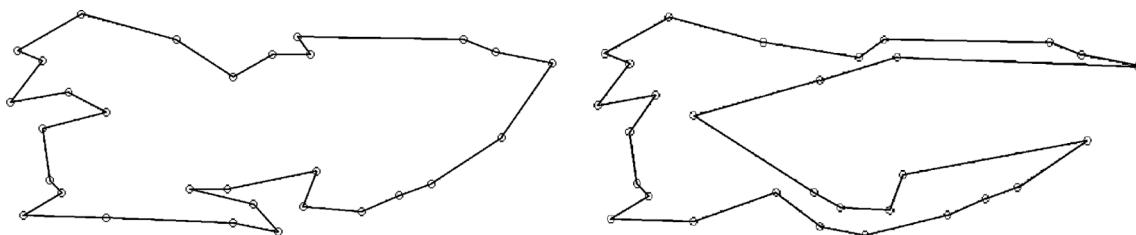


Fig. 1. Two solutions to the same ETSP stimulus (generated using R). The left solution has a surface area of 5213 AU^2 , while the right solution has a surface area of 2817 AU^2 . Participants will typically judge the left solution as being more optimal. Note. AU is arbitrary units and depends on the screen resolution, but the ratio between the area of the solution and the total area remain consistent.

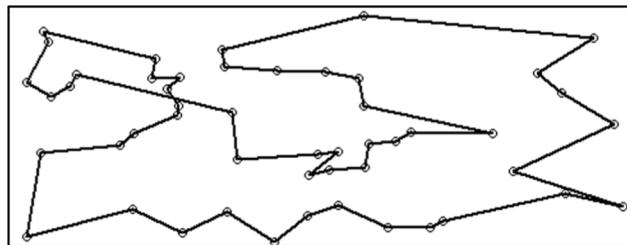


Fig. 2. Solution to a 50 node ETSP containing a single crossing.

increase in processing time, as a result of number of nodes, indicates that the production of ETSP solutions are not determined by local processing (System 2), but as a result of heuristic/global processing (System 1). Human solutions can, at times, even outperform computational heuristics (Kyritsis et al., 2018b), making it pertinent to model human behaviour in order to improve and/or develop faster algorithms that produce near-to-optimal ETSP solutions.

Existing research has identified that human solutions tend to be both void of edge crossings, and tend to display convexity, which has resulted in researchers proposing a plethora of behavioural models such as: global-to-local precedence (e.g., convex hull hypothesis, and the quotient algorithm), local-to-global precedence (e.g., crossing-avoidance hypothesis and pyramidal models), or a mix of both (e.g., generative models) (Kyritsis et al., 2018b; MacGregor & Chu, 2011; MacGregor, Chronicle, & Ormerod, 2004; Pizlo et al., 2006; Van Rooij, Stege, & Schactman, 2003). Despite a differing of opinion concerning which cognitive approach is dominant, there is consensus that several geometric properties, termed ‘figural effects’, impact human ETSP solution generation performance, e.g., indentations, number of edge crossings, and convexity (Kyritsis, Gulliver, & Feredoes, 2018a; MacGregor & Ormerod, 1996; van Rooij, Schactman, Kadlec, & Stege, 2006). An indentation is a notch on the edge or surface of the route convex hull. MacGregor (2012) investigated whether the number of indentations influenced participant perception of optimality in a discrimination task, i.e., when different TSP solutions were presented in pairs. The results of MacGregor (2012) study indicated that, even though participant solutions tend to have fewer number of indentations, participants did not use number of indentations when comparing optimality. Accordingly, there appears to be a discrepancy between the geometric properties and processes used to generate solutions, and those used to judge the optimality of existing tours. Nevertheless, there is very limited research aimed at understanding the figural effects that impact judgement of optimality (tour lengths) of existing tours, i.e., discriminating between different pre-existing solutions (comparing the efficiency of existing solutions for a specific node layout). In this research we hypothesise that extracted geometric properties have an effect on stimulus selection in a discrimination task (selection or no selection).

Ormerod and Chronicle (1999) found a linear relationship between aesthetic perception of a ETSP solution and the ratio between internal nodes, i.e. nodes not included in the convex hull and the number of

nodes included in the convex hull. This result implies that participants visually prefer ETSP solutions where most nodes sit on the convex hull (e.g. implying that solution shown in Fig. 1 left is preferred over the solution in Fig. 1 right). In contrast, however, Vickers, Lee, Dry, Hughes, and McMahon (2006), whilst controlling for regularity, number of potential intersections, number of internal nodes, and the number of nearest neighbours in each solution, failed to find evidence that the convex hull is an important determinant of aesthetic appeal.

Dry and Fontaine (2014) focused on discrimination of tour lengths, and asked participants to choose the easiest ETSP graph to solve from four competing node layouts. The authors concluded that the number of convex hull nodes and the number of possible intersections served as the selection criteria for the discrimination test. This study explored selection between stimuli with a range of inherent geometric properties – particularly number of possible intersections, and the number of convex hull nodes. Although helpful in understanding ETSP solution formation, Dry and Fontaine (2014) did not consider participants’ ability to compare pre-existing solutions, but instead requested participants to select from a set of inconsistent unsolved layouts. MacGregor (2017) investigated the impact of two geometric properties, i.e., number of crossings (complexity) and area (convexity), on judgement of optimality between competing solutions. The author concluded that edge crossings, and the area of the polygon were significant to the participants’ judgement of optimality. Ultimately, solutions with less crossings, and solutions with larger polygon areas, were considered to be more optimal, with edge crossings the most important of the two criteria. Interestingly, the presence of any crossings in a solution, as illustrated in Fig. 2, renders the solution, from the view of a participant, as inherently ‘sub-optimal’. In the absence of crossings, however, polygon-enclosed area is used as the primary selection criteria, i.e. with a larger convex area representing a more optimal solution.

Since some geometric properties have been shown to correlate (Vickers et al., 2006), it makes sense to compare stimuli as pairs, or groups of solutions, as this can decrease finding uncertainty as a result of unknown and/or uncontrolled variables. For this study, we were interested in exploring the impact of the geometric properties (i.e., regularity, area, tour cost, and path complexity) on judgement of ETSP optimality and ability to discriminate optimality. To satisfy this aim, a data mining and modelling approach was used to extract geometric properties from thousands of participant trials (independent variables) in order to better understand how they affect the process of target selection in a discrimination task (dependent variable), where participants were asked to decide between competing ETSP stimuli based on perceived optimality.

2. Method

2.1. Experimental procedure

The experimental front-end platform was developed using HTML5 and Javascript, whilst PHP and mySQL were used to code the back-end processes and data storage. The platform was tested on Google Chrome,

Table 1

Description of heuristics used to produce the four solutions in the experiment.

Heuristic	Description
Nearest Neighbour (NN)	Find the closest unconnected node and join this node to the sub-tour. Repeat until all nodes join the sub-tour to form the final tour.
Nearest Insertion (NI)	Join the closest node to form a sub-tour. Find a new node not in the sub-tour that is closest to any of the sub-tour nodes. Insert the new node between two nodes of the sub-tour in such a way as to minimise the cost. Repeat until all nodes are inserted.
Farthest Insertion (FI)	Join the farthest node to form a sub-tour. Find a new node not in the sub-tour that is farthest from any of the sub-tour nodes. Insert the new node between two nodes of the sub-tour in such a way as to minimise the cost. Repeat until all nodes are inserted.
Cheapest insertion (CI)	Join the closest node to form a sub-tour. Find the node not in the sub-tour that causes the least amount of increase in the cost. Insert the new node between two nodes of the sub-tour. Repeat until all nodes are inserted.

Microsoft Edge, and Mozilla Firefox. Data collection was anonymous, i.e., no IPs or contact details were collected. However, to stop participants from repeating the experiment more than once, PHP sessions were used to stop participants who (perhaps accidentally) tried to run the experiment multiple times. To allow collection of anonymous data, a unique number was generated and assigned to each participant. Randomly generated ETSP graphs were created using R by generating X and Y coordinate sets from a uniform distribution ranging from 0 to 100, such that $X = X_1, \dots, X_n \sim U(0, 100)$ and $Y = Y_1, \dots, Y_n \sim U(0, 100)$. The set of samples were then combining to create graph nodes G , so that $G_1 \dots G_n = (X_1, Y_1) \dots (X_n, Y_n)$ and rounded off G so that $G \in \mathbb{N}^2$. The number of nodes (n) ranged from 30 nodes to 70 nodes (in 10 node steps). Ultimately, 2,000 ETSP graphs were generated for each step increment (i.e., 2,000 30 node graphs, 2,000 40 node graphs, etc.), making 10,000 graphs in total. For each of the 10,000 graphs, with the help of the TSP library (Hahsler & Hornik, 2007), solutions were generated by applying four commonly used TSP solution heuristics: Nearest Neighbour; Nearest Insertion; Cheapest Insertion; and Farthest Insertion (see Table 1 for details of the four applied heuristics). The starting node was kept constant for all heuristics. The stimuli presented to participants was rendered on four separate HTML5 canvases (see Fig. 3). The solutions were assigned to a cell position randomly, i.e., solutions generated by

each of the heuristics had an equal likelihood of being assigned to each cell. A mouse listener was implemented to record participant mouse clicks as a method of logging target selection within each trial. Time-stamps were also coded to record response time (RT); where $RT = Time_{end} - Time_{start}$.

Upon providing consent, i.e., by clicking the ‘happy to provide consent’ checkbox, a short tutorial appeared on the screen explaining the travelling salesperson problem and the experimental process. This was done using a combination of step-by-step descriptions and illustrations across two separate pages. For the sake of transparency, we have added the instructions in Appendix 3. Participants were made aware that all trials were timed and were asked to complete trials as quickly as possible; without compromising the quality of their choices. Once the participants were ready to start, they were asked to click a hyperlink labelled ‘Start Experiment’ at the bottom of the page.

Before each trial, a black cross was presented for two seconds in the middle of a grey screen, i.e., to maximise the chance that all participants started at the same gaze position. The cross was followed by presentation of the four possible solutions (see Fig. 3). Participants were instructed to click on the solution that they deemed to be the most optimal – in terms of tour length, i.e., the least distance ‘travelled’ by the salesperson. The starting/finishing node was highlighted in green on all solutions throughout the experiment. Ten randomly selected trials were presented from each of the five increments, i.e., 50 trials per participant. Upon completion of the experiment, participants were directed to a ‘Thank you’ page, which asked them to close their browser tab. The entire procedure is summarised in Fig. 4.

2.2. Participant sample

The proposed experimental procedure received a favourable ethics review in accordance with the ethical standards of the School of Psychology and Clinical Language Sciences, at the University of Reading (UK). 120 participants were recruited using simple random sampling through email and social media. Only participants that had normal or corrected to normal vision were invited to participate in the experimental study. All subjects were explicitly made aware that involvement in the study was voluntary and that they could withdraw at any time by simply closing the browser tab. Consent forms and information sheets were converted to HTML and all participants were asked to confirm that they were happy to participate in the experimental study prior to sitting

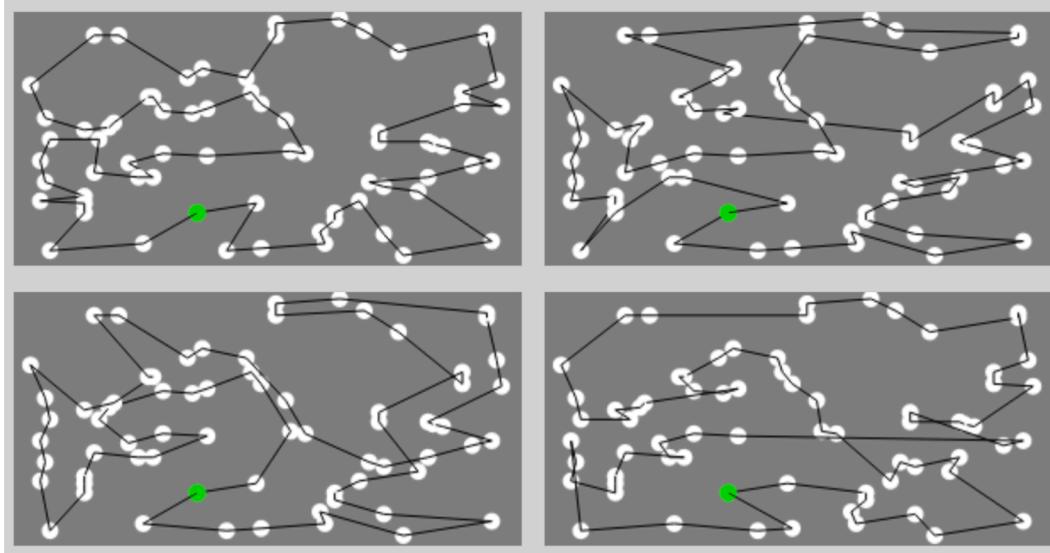


Fig. 3. Four heuristic solutions to the same randomly generated ETSP graph were coded to appear in each trial. The green shaded node is the start/finish node. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

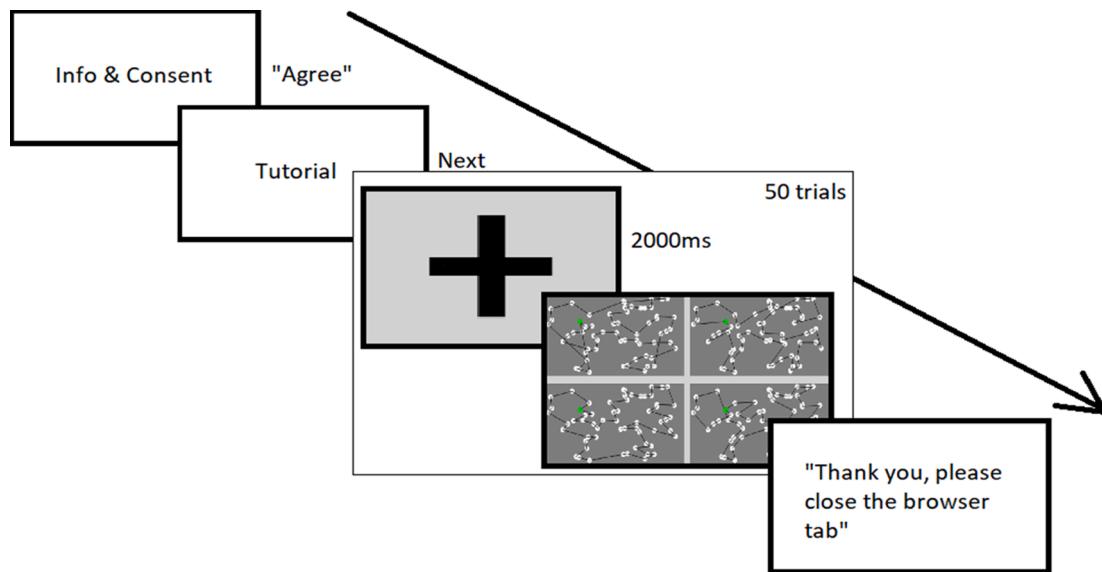


Fig. 4. Summary of the experimental procedure.

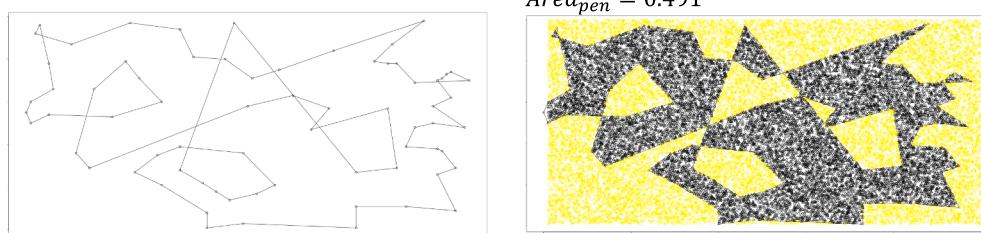


Fig. 5. Graph contains internal edge crossings (left). The Monte Carlo algorithm will penalize these graphs by ignoring the areas created internally (right) Note. The area formed within the polygon-enclosed area is not considered when using the Monte Carlo approach (described in [Appendix 1](#)). We suggest that these inner intersections mirror the lack of convexity in these solutions.

Table 2
Description of geometric properties extracted from the TSP solutions.

Properties	Method
$Area_{pen}$ ^a	The polygon enclosed area of the solution with a penalty for internal crossings
Complexity	Number of edge crossings
Regularity	Variation (SD) of the length of the edges in the solution
Cost	The total tour cost (i.e., the sum of length of all edges that make up the solution)

^a The area is an approximation that was extracted using Monte Carlo methods, which introduce a penalty for internal crossings.

through the experiment. A script ensured that participants had ticked the box before allowing them to continue. Gender, age, and other personal data were not recorded for this study. No compensation was provided for participating in this study.

2.3. Extraction of figural effects

All data was analysed using R (R Core Team, 2016), the ‘BayesFactor’ library (Morey, Rouder, & Jamil, 2015), and the ‘caret’ package (Kuhn et al., 2018). We modelled the discrimination task using selection vs non-selection, as well as success vs failure (i.e., with success defined as the selection of the solution with the most optimal solution) as the two response variables, and the extracted figural effects as the predictor variables. Several studies (e.g., Ormerod & Chronicle, 1999; Vickers et al., 2006; Dry & Fontaine, 2014; MacGregor, 2017) suggested ways to

extract tour convexity, tour complexity, and tour circularity; shown to impact human problem-solving when generating solutions for the ETSP. The measure of circularity is closely related to ‘compactness’, i.e. the area of the polygon over its perimeter, which has been shown to influence aesthetic preference for polygons (Friedenberg and Bertamini, 2015). Therefore, circularity can be extracted quite easily in graphs with no internal crossings (i.e., crossings that occur by routes going back inside the polygon). When considering convexity, we suggest that graphs with internal crossing should be penalised, because crossings can result in areas inside the convex hull that are not considered part of the final graph solution (e.g., [Fig. 5](#)). Accordingly, instead of using the term ‘convexity’ for this measurement, we denote this measurement to be $Area_{pen}$, since convexity is measured by taking A_s/A_c (where A_s is the polygon-enclosed area of the solution and A_c is the area of the convex hull surrounding the solution).

[Table 2](#) summarises the geometric properties that were extracted from the heuristic solutions. Note that since the ETSP stimuli were kept consistent in our experiment, we use the variability in the length of edges as an approximate measure of regularity, rather than the clustering of stimulus (as presented by Dry & Fontaine, 2014). By definition, an irregular polygon will have both unequal edge lengths and angles, however, because of the uncertainty of what constitutes an ‘internal angle’ for intersecting polygons, we did not consider this measurement. A full description of how these geometric properties were extracted from the graphs has been added to [Appendix 1](#).

In our experimental set, to apply a penalty to solutions that contain internal crossings, we did not consider the internal area created by crossings (as shown in [Fig. 5](#)). It is important to clarify that this is a

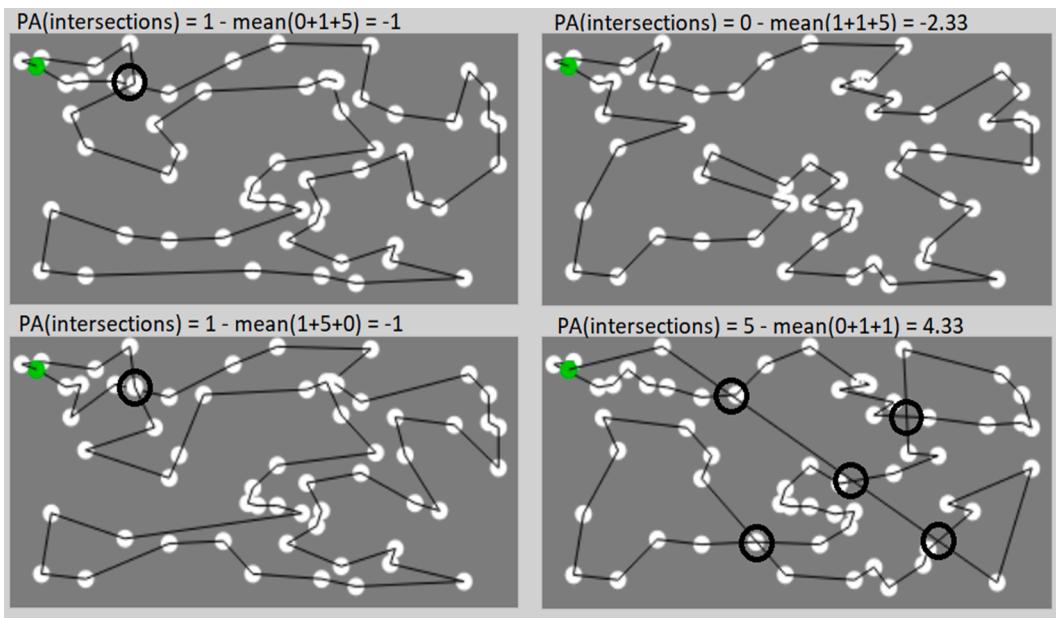


Fig. 6. Crossings are highlighted (i.e., top-left graph has one crossing, and bottom-right has five). For each stimulus, perceptual activation for a geometric property is derived by subtracting the quantity of said property in the stimulus by the average quantity of its neighbours.

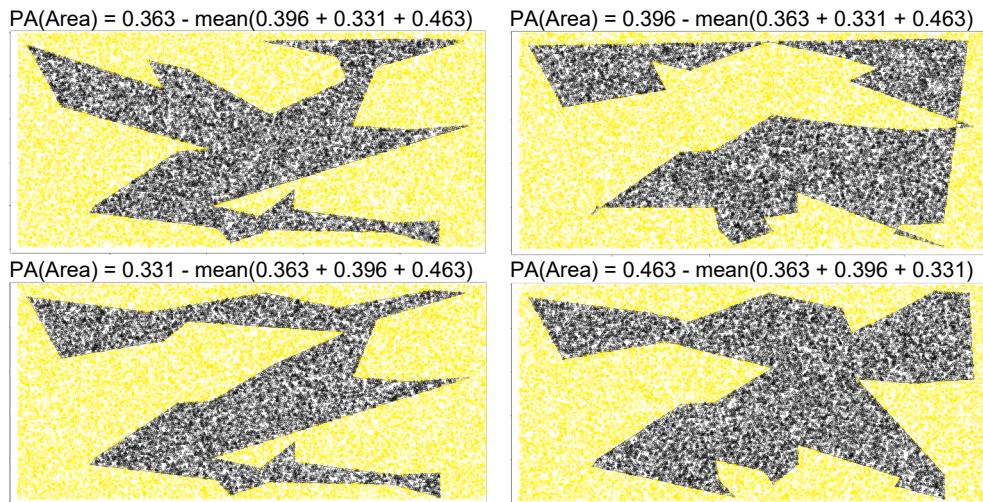


Fig. 7. Perceptual area activation scores. Note that the top-right stimulus contains crossings, yet has a larger area than two of the competitors with no crossings.

numerical method that penalizes irregular solutions that contain internal crossings, and are therefore difficult to measure in terms of convexity. We are not suggesting that the human visual system processes the information in this way. Rather, we suggest that the presence of internal crossings deters participants from considering any solution that contains internal crossings as a possible candidate in the discrimination task; limiting perceptual consideration of area.

2.4. Perceptual activation of properties in the discrimination task

Each trial contained four measurements for each of these geometric properties, i.e.: area, tour costs, complexity (number of edge crossings), and regularity (standard deviation of the length of edges in solutions). Since studies suggest that humans approach the Euclidean TSP using global processing, we presume that properties between competing solutions (stimuli) are computed at a perceptual level, i.e., in order to support decision-making in near-linear time. Hence, rather than using absolute measurements of each stimulus separately, the difference in the

quantity of geometric properties between each stimulus and its surrounding neighbours (for each trial) was extracted using the formula:

$$P_{inter} = \omega_1 - \frac{1}{3} \sum_{i=2}^4 \omega_i. \quad (i)$$

where ω_1 is the score of the geometric property of any of the four graphs, and $\omega_2, \omega_3, \omega_4$ are the geometric scores of the remaining graphs in the trial (i.e., its competitors) – see Fig. 6.

In Fig. 6, the top left stimulus has one intersection. Therefore, the ‘perceptual activation score’ for complexity would be one minus the average number of intersections in the other graphs, giving it the number negative one. This approach allows us to make sense of trials where complexity is equal for all four competing stimuli, i.e., they all have the same number of crossings, and therefore complexity cannot be used as a guidance property. Note, that for regularity, complexity, and tour costs, perceptual activation scores decrease with optimality. Area perceptual activation scores increase as optimality increases (see Fig. 7).

Table 3

Model comparisons of four logistic models. BIC was used as the criterion for retention of variables that predict stimulus selection.

Model no.	Perceptual Activators	BIC	Bayes Factors	R ² m	R ² c
1.	+ pArea _{pen}	18742.06	3.96 × 10 ²²⁷	0.04	0.10
2.	Model 1 + pComplexity	17806.66	3.00 × 10 ²⁴	0.13	0.19
3.	Model 2 + pCost	17693.94	1	0.14	0.20
4.	Model 3 + pRegularity (full model)	17703.49	118.48	0.14	0.20

Note. Bayes Factors were taken in relation to the most likely fit (i.e., the lowest BIC score), as shown by Dry et al. (2006). Model 3 was the best fitting model (highlighted).

Table 4

Standardised coefficients and z values for the logistic mixed effects model.

Perceptual Activators	Estimate	Std. Error	Z value
Cost	-0.007	<0.001	-10.97***
Complexity	-0.134	0.011	-11.85***
Area _{pen}	1.324	0.309	4.29***

***p < 0.001.

Table 5

Correlation table of fixed effects of the full model (all variables).

	Intercept	pArea	pComplexity
pArea	0.045		
pComplexity	0.008	-0.504	
pCost	0.104	0.769	-0.681

3. Analysis

3.1. Human performance in the discrimination task

The point-estimated probability of a participant selecting the most optimal route was 0.47 [Bootstrapped 95% Confidence Intervals from 5000 samples: 0.43 < μ < 0.51]. Interestingly, existing work in the field (see MacGregor, 2017) placed the probability of successfully discriminating competing tours as being higher – based on complexity (i.e., crossings) at 66%, and on area at 91%. It seems unwise to ignore or dismiss these existing findings. Accordingly, we used a Bayesian binomial analysis and took the rounded mean of the two readings as the maximum likelihood estimate of a beta prior, i.e., $\theta \sim Beta(16, 4)$. In turn, we calculated the likelihood of occurrence, i.e., $p(D|\theta)$; and the combined knowledge was used to generate a posterior distribution using Markov Chain Monte Carlo (2 chains, 20,000 iterations), i.e., $p(\theta|D) \propto p(D|\theta)p(\theta)$. Our results suggest that the maximum likelihood estimate for the posterior $\hat{\theta} = 0.54$, with a 95% credible interval ranging from 0.46 to 0.63. Regardless, our findings are much less optimistic than findings from other studies.

Existing studies propose that several figural effects from Euclidean graphs impact human performance when generating solutions to the ETSP. These figural effect include: distance and variability between the nodes that form the convex hull of the graphs; the number of nodes that make up the convex hull; number of potential intersections in the graph; clustering of the graphs; and the variability of the angles of the convex hull (Dry, Preiss, & Wagemans, 2012; Kyritsis et al., 2018; Vickers, Lee, Dry, & Hughes, 2003). MANOVA with the aforementioned figural effects set as dependent variables and success set as the independent variable, which is a binomial variable – i.e., yes/no, indicated that there was an overall significant difference in some of these properties when comparing successful and failed trials [$F(6, 4223) = 9.66$, $p < 0.001$, $V = 0.01$]. Independent ANOVAs, followed by Tukey HSD as a post-hoc test

for multiple comparisons indicated that the following measures were significantly higher in successful vs failed trials: distance of nodes in the convex hull of the graph ($M_{diff} = 1.34$, CI 95% [0.35, 2.35], $p < 0.001$), variability (standard deviation) of angles in the convex hull ($M_{diff} = 1.13$, CI 95% [0.50, 1.77], $p < 0.001$), and clustering of the graph ($M_{diff} = 3.03$, CI 95% [2.03, 4.04], $p < 0.001$). However, it is worth noting that in all cases the effect size difference, calculated using Cohen's d , was negligible: $d = 0.08$, $d = 0.11$, and $d = 0.18$, respectively. Therefore, we found evidence that suggests that some of the figural effects that have been previously shown to impact human performance when generating TSP solutions in Euclidean space, also have a negligible effect on human performance when discriminating between existing solutions.

3.2. Likelihood of selection as a function of perceptual activation scores

In every trial, participants selected one graph that they perceived to be the most optimal TSP solution. Regardless of whether the selection was 'correct', i.e., was indeed the most optimal graph, we expect that the perceptual activation scores of the figural effects considered in this study influenced that choice. Stimulus selection can be thought of as a binomial dependent variable. Accordingly, we examined the likelihood of a specific stimulus being either selected (or not) as a function of the weighted influence of its associated perceptual activation scores. Since the experiment relied on a repeated measures design, a generalised linear mixed model was implemented by applying the 'lme4' package (Bates, Mächler, Bolker, & Walker, 2014) in the form:

$$\text{logit}(\text{selection}) = \beta_1 * pArea + \beta_2 * pComplexity + \beta_3 * pRegularity * \\ + \beta_4 * pCost + (tID|ID), \quad (\text{ii})$$

where ID is the random effects variable for participant ID, tID is the trial ID that varied for each participant, and $pArea$, $pComplexity$, $pRegularity$, and $pCost$ are the perceptual activation scores for respectively area_{pen}, number of edge crossings, standard deviation of edges, and tour costs. For model comparisons, predictors from the model were removed in descending order of standardised coefficients. To evaluate how much the model suffered at each step, we compared the delta of BIC (Bayesian Information Criterion) to the best fitting model. This in turn allowed us to evaluate model fits using Bayes Factors analysis, which Wagenmakers (2007) suggested could be extracted from ΔBIC using:

$$BF_{10} = \exp((BIC2 - BIC1)/2). \quad (\text{iii})$$

Comparisons of each model, including the predictors, BIC scores, Bayes Factors, marginal R^2 , and conditional R^2 can be found in Table 3.

Marginal and Conditional R^2 were extracted with the help of the 'MuMin' package (Barton, 2020). Frequentist analysis suggested that graph volume (i.e., the number of nodes) significantly impacts the chance for successfully selecting the most optimal target, however the effect size was small [$\chi^2(4) = 22.58$, $p < 0.001$]. The independent multinomial Bayes Factor analysis resulted in $BF_{10} = 1.19$, which, according to Kass and Raftery (1995), is not sufficient evidence to support independence of the groups; i.e., there does not appear to be an association between graph volume and the chance of successfully selecting the optimal solution.

Model comparisons using Bayes Factors indicates that Model three, which uses the perceptual activation scores of area, complexity, and cost (see Table 3) was the best fitting model. The next best model (i.e., four) included consideration of regularity, was ~ 118 times less likely to fit the data. According to Kass and Raftery (1995) our result is decisive evidence in favour of model three, and provides credible evidence that all perceptual activation scores, with the exception of regularity, played a key part in driving stimulus selection.

The standardised coefficients (a measure of variable importance), the z values, and the p-values of the best fitting model can be found in Table 4, which show that area, when considering the penalty term for internal edge crossings, is more meaningful in determining likelihood of

Table 6

Description of the three main parameters returned by drift–diffusion models.

Parameter	Description
v	The mean drift rate that influences the direction that a process would take towards boundary A or B
a	The distance between the thresholds of the decision boundaries of two choices: A and B
z	The starting point of the process. This is indicative of bias towards one decision over the other.

Note. For a visual illustration of how the parameters affect decisions in our choice discrimination task see Fig. 8.

Table 7

Conditions used as inputs in the drift diffusion model.

Condition	Favourable Perceptual Activation
Area	$P_A(\text{Area}) > 0$
Complexity	$P_A(\text{Complexity}) < 0$
Cost	$P_A(\text{Cost}) < 0$
Regularity	$P_A(\text{Regularity}) < 0$

Note. The function P_A is the perceptual activation function as shown in Figs. 6 and 7.

Table 8

Parameters indicate the average starting bias, drift-rate, and selection boundary width for selection vs non-selection in our discrimination task according to the D*M drift–diffusion model.

Parameter	α	v	z
Intercept only (RT)	0.33	−3.76	0.66
Complexity	1.9	0.27	0.29
Area	2	0.9	0.5
Cost	2	0.23	0.3
Regularity	2	−0.27	0.45

Note. Regularity has a negative slope, which implies that it is not a strong enough perceptual activator to drag away from the default diffusion towards non-selection (as shown in the intercept term).

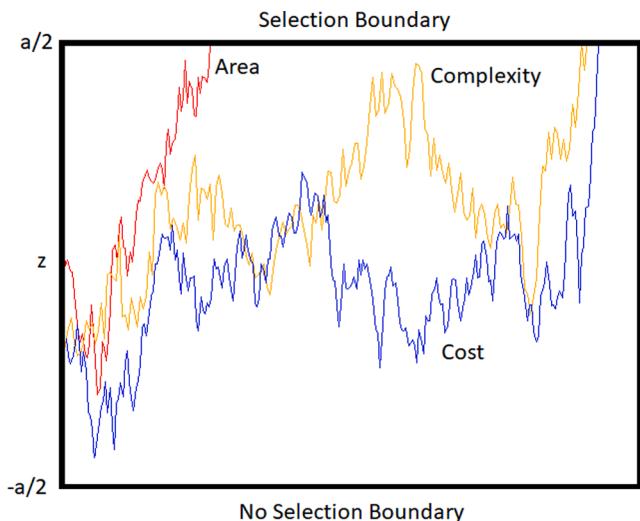


Fig. 8. Random walk simulation using the average parameters of the drift–diffusion model. Note. The drift-rate for the area perceptual activation score is much steeper than the other two activators, further validating our mixed-effects model, that showed perceptual score of area, with the associated penalty term, to have the strongest effect on target selection.

Table 9

Regression modelling using BIC as the criterion for retention of variables that predict the likelihood of a successful trial.

Model no.	Perceptual Activators	BIC	Bayes Factors	Pseudo- R^2
1.	+ pArea	1833.32	1.04×10^{199}	0.61
2.	Model 1 + pRegularity	931.45	1509.56	0.81
3.	Model 2 + pComplexity	916.82	1	0.81

Note. Bayes Factors were taken in relation to the most likely fit (i.e., the lowest BIC score), as shown by Dry et al. (2006). In this case, model two is 1510 times less likely to fit the data than model three.

selection than edge crossings alone.

Table 3 shows that fixed effects accounted for 14% of the variance in the stimulus selection, i.e., the marginal variance R^2m , while random effects accounted for an extra 6% of the variance. The model, however, suffered from correlation issues (see Table 5), yet the model did not suffer from major multicollinearity issues (as all VIF scores were under 5). This result indicated that the perceptual activation scores were at the very least not linear combinations of each other, and, therefore, showed fair independence.

The logistic model provides evidence to partially support our initial hypothesis, i.e., that the perceptual activation scores of the extracted geometric properties have an effect on stimulus selection in a discrimination task (selection or no selection). However, the model is not very informative with respect to the process of making a choice (selection or no selection), and how reaction time (RT) is affected by competing perceptual activation scores.

3.3. Modelling selection and reaction time using a drift-diffusion model

Selection process, and the impact on RT, can be considered using diffusion models. A diffusion model belongs to the family of sequential-sampling models, and is often applied to the analysis of discrimination tasks that contains two or more choices – for a comprehensive review of diffusion models see Ratcliff and Smith (2004) or Wagenmakers (2009). Diffusion model theory states that as participants are presented with stimuli (cues), they accumulate the cues presented to them, process the cues, and ultimately make a choice from selections A or B (Alexandrowicz, 2020). The process itself, however, is noisy and stochastic, often simulated as a random-walk, with drift parameters rather than constant slope terms (i.e., drift–diffusion models). This is reasonable as (i) cognitive tasks are not guaranteed to lead to success – even in the most informative environments, (ii) decisions may vary, even for trials with the same conditions – and even in repetitions by the same participant, and (iii) there is variation in the time taken to complete the task – even when participants repeat the same task (Smith & Ratcliff, 2015). Drift-diffusion models return several parameters, which help the process of decision-making to be understood. The main parameters that we consider in the analysis are discussed in more detail by Voss, Rothmund, and Voss (2004) and are summarised in Table 6.

The unstructured nature of our experimental design meant that reverse consideration of trials was required, i.e., identification of trials where only one perceptual activation score differed significantly compared to their competitors. As there was not enough information about the competitor stimuli to allow comparisons between them, we consider the outcome of each trial as being binomial: either a graph was selected by the participant or it was not. In the end, we isolated several trials that allowed consideration of conditions as shown in Table 7.

Next, to ignore non-decision time, which was not factored into our experimental design – see Verdonck and Tuerlinckx (2016) for a description, we applied the D*M drift–diffusion model method; applying the 'DstarM' package in R (van den Bergh, Tuerlinckx, & Verdonck, 2020) to extract the parameters highlighted in Table 6. To support model building, the data was split to ensure balance, i.e., resulting in

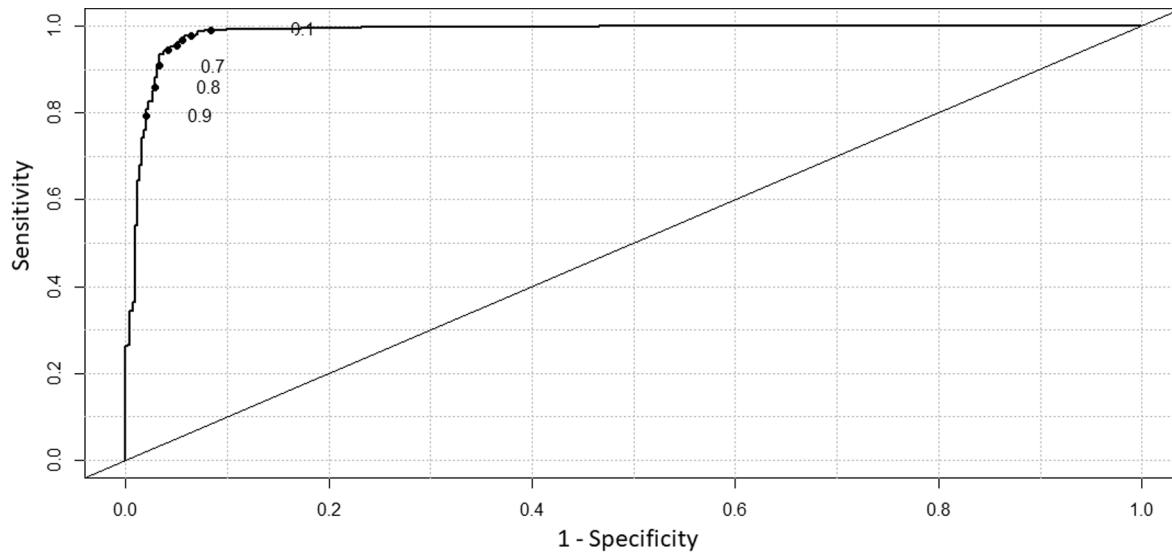


Fig. 9. The Receiver Operating Characteristic curve highlights the best trade-off between sensitivity and specificity (which is the cut-off at the top-left corner of the curve).

Table 10

Confusion matrix of the predictive logit model shown in (v) with a cut-off of 0.7.

Reference			
Prediction			Success
	Fail	Success	
Negative	438	35	
Positive	15	358	

consideration of only (i) the selected stimuli, and (ii) an equal number of non-selected stimuli (i.e., one of the non-selected stimuli from each trial). Moreover, only trials that lasted less than seven seconds were considered, as this delay is in line with expected results from a global processing task. In total there were 2114 trials. The parameters for each condition is summarised in Table 8. Fig. 8 shows, for the sake of demonstration, a hypothetical trial using the average slopes of each condition in a random walk simulation.

Both D*M drift-diffusion and logit models show that the effect on selection follows the order: Area_{pen}, Cost, and then Complexity. As such, the models have illustrated two things: firstly, that perceptual activation scores of geometric properties work as weighted inputs of a function that returns the likelihood of stimulus selection in the discrimination task, and secondly, that perception of area, with the associated penalty term for internal crossings, seems to be particularly influential in guiding human selection.

3.4. Likelihood of success as a function of perceptual activation scores

To validate our extracted properties as drivers for determining optimality in the discrimination task, we used a predictive analytics approach. It is already well established that the most optimal solution will never contain edge crossings, however, less is known about the properties of regularity and the area of the enclosed polygons. In this case, it would not make sense to include tour costs in our model, since the most optimal solution will have the lowest tour costs. A logistic model was fit, i.e., to evaluate whether perceptual scores can make any meaningful prediction of the likelihood of a stimulus being successful (note that we did not account for random effects, since no generalisations about human behaviour were made in this section).

$$\text{logit}(\text{success}) = \beta_0 + \beta_1 * p\text{Area} + \beta_2 * p\text{Complexity} + \beta_3 * p\text{Regularity}. \quad (\text{iv})$$

The results of the full model show that all three perceptual activation scores were significant predictors of the likelihood of a stimulus being

successful (see Table 9 for model comparisons). As before, variables were removed in order of least importance from the full model, i.e., according to their standardised coefficient scores.

We were interested in evaluating how this model would perform on unseen data. To provide a proof of concept, the dataset was split into a training and a test set (80/20 split) – with the help of the ‘createDataPartition()’ function found in the ‘caret’ library (Kuhn et al., 2018) – and was used with the full model to predict the probability of unseen trials being classified as successful. The Receiver Operating Characteristic (ROC) curve (shown in Fig. 9) illustrates that a suitable cut-off (trade-off between sensitivity and specificity) is ~ 0.7 .

The ROC curve indicated that the probability of an observation being classified as a success is higher than 0.7. More formally,

$$\hat{y}_i = \begin{cases} 1 & \text{if } P(\hat{y}_i) > 0.7 \\ 0 & \text{otherwise} \end{cases} \quad (\text{v})$$

where \hat{y}_i are the fitted probabilities of the logit model on previously unseen data (i.e., the predictions on the test set). Applying the cut-off of 0.7 – as shown in (v) – leads to the confusion matrix found in Table 10.

The most interesting finding from this confusion matrix is the very high positive predictive value, which is the probability of a successful trial given a positive prediction, i.e.,

$$P(\text{Success} | \text{Positive}) = \frac{P(\text{Success AND Positive})}{P(\text{Positive})} = \frac{358}{373} \approx 0.96 \quad (\text{vi})$$

The result shows that, using just these three variables, our model can accurately predict 96% of the time, whether a stimulus is likely to be the most optimal solution. Interestingly, however, participants did not perform as well as the regression model in the task, which suggests there are other factors influencing humans when selecting the most optimal from multiple pre-defined ETSP solutions.

4. Discussion

In this study we presented ETSP solutions (with node volume between 30 and 70) to participants in an online experiment. Participants were asked to select the one most optimum solution, i.e., the stimulus with the shortest tour length in a discrimination task between four competing stimuli (ETSP solutions). Solutions were produced using common TSP heuristics (Nearest Neighbour; Nearest Insertion; Cheapest Insertion; and Farthest Insertion) and were randomly positioned on a 2x2 grid. Data mining was used to extract geometric properties, which

literature suggests would aid human selection in an ETSP solution discrimination task, and aid human selection response time. The extracted properties were complexity (edge crossings), polygon area with an associated penalty term for internal crossings, regularity (SD of edge lengths), and tour costs. Complexity was extracted from stimuli through a system of two linear equations (see xix in [Appendix 1](#)). Area of the enclosed polygon was extracted using the Monte Carlo method (see formulas vi to xvi in [Appendix 1](#)). Regularity and tour costs were extracted by measuring edge lengths, and in the case of regularity the standard deviation of all the lengths of edges in a solution (see formula xxvi in [Appendix 1](#)). Finally, the properties were used as inputs in a perceptual activation function, which returned the difference between the properties in competing stimuli, rather than treating them as absolute values. Admittedly, one of the limitations of our study was that only a small set of properties that impacted target selection were investigated, most of which had been previously discussed in literature (with the exception of regularity); i.e., we are not presenting an exhaustive list of variables.

Participants performed quite poorly in this task (point estimate of 47% success rate), and success was only negligibly influenced by the usual properties often shown to occur in the formation of participant solutions (such as convex hull parameters). Poor participant performance contrasts with the more optimistic results found by [MacGregor \(2017\)](#), however, the difference in results were speculative due to (a) the noisy nature of online experiments, and (b) that our competing solutions were generated on the same graph using four different heuristics, none of which are ‘state of the art’. It is therefore possible that participants performed more poorly than anticipated due to all solutions simply being too underwhelming.

Mixed-effects linear modelling was used to evaluate whether the probability of stimulus selection can be expressed as a function of perceptual activation scores. Our findings suggest that differences in polygon area, complexity, and tour costs explain a small, but significant amount of the variance in participant selections in the discrimination task, with about half of that attributed to conditional / random effects. Regularity did not predict likelihood of selection, despite being a significant predictor of optimality when used as an input in a logit classification model. Furthermore, the standardised effect sizes indicate that the perceptual activation score from the polygon area measurement, when accompanied by a penalty term for overlapping areas caused by internal crossings, was the most important cue for target selection, followed by complexity, and then tour costs. This finding was reinforced by our drift-diffusion model, which also showed a larger drift-rate for area, followed by complexity, and finally cost (see [Fig. 8](#) for the output of a random walk simulation given the parameters of the drift-diffusion model). It is worth noting that we were forced to set a cut-off *RT* for this model at seven seconds to be more in-line with what one would expect from a global processing task. This resulted in a loss of about half of the trials and a significant drop, therefore, in statistical power. However, the drift-diffusion model is an early theoretical framework for modelling the process involved in reaching a decision (selection between competing stimuli) boundary in the task through accumulation of evidence (i.e., perceptual activation scores); though this needs to be validated in a lab setting whilst testing for other properties that impact judgement of optimality in this discrimination task.

As mentioned earlier, the area of graphs in our study, when using the Monte Carlo method, was penalised by the presence of internal crossings (i.e., crossings that occur by routes that go back inside the polygon itself, as shown in [Fig. 5](#)). Our data suggests that people are averse to selecting these types of graphs when judging optimality, more so than other solutions that contain external crossings. As an ad hoc follow-up to these findings, we asked ten participants to judge two solutions, one with internal crossings, one with external crossings, while consistently maintaining all other properties. The emerging codings identified that solutions with internal crossings were perceived as being “messy”, “intertwining”, “back-and-forth”, and “looping”. As expected, all

participants described solutions with internal crossings as being less optimal. Our anecdotal evidence suggests that, despite the presence of crossings being a well-known factor that impacts human performance in the TSP, for the discrimination task this can be further broken down into internal and external crossings. From a perceptual viewpoint, we believe the presence of internal crossings more strongly activates the diffusion process towards the non-selection boundary than external crossings. We aim to explore the reason for this further in future work.

Our results are consistent with the work of [MacGregor \(2017\)](#), who argued that both complexity (edge crossings) and area are important cues for discriminating between ETSP tour lengths. Furthermore, we expanded the area measurement to include a penalty for internal crossings as to not over-bias the estimated area that these solutions would otherwise create.

Moreover, our results agree with previous findings, which suggest that geometric properties in the ETSP task correlate with each other ([Vickers et al., 2006](#)), which, as [MacGregor \(2007\)](#) stated “suggests a nexus of stimulus characteristics that people could use to judge the lengths of tours”. Our correlation matrix ([Table 5](#)) shows that properties were significantly correlated; however, this is practically not surprising when altering one will subsequently alter the others. It is important to note that our logistic model did not indicate serious multicollinearity problems (all VIFs < 5), which at least leads us to believe that the perceptual activation scores were fairly independent, and not just linear combinations of each other.

Finally, a logit model, with success as the binomial response variable and perceptual activation scores as the predictors can be used to detect the most optimal solution 96% of the time in the discrimination task, which shows that extracted geometric properties (i.e., area, complexity, and regularity) can be used effectively to distinguish between ETSP solutions – however there are other factors impacting human trials.

5. Conclusion

Studies have consistently shown that some humans can outperform even the best commonly used heuristics when solving ETSP graphs (see [Kyritsis et al., 2018b](#)). Although humans are good at ETSP problem solving, they performed poorly in a discrimination task where they were asked to judge between four competing ETSP stimuli and select the most optimal solution. Our initial probability of success (*p*(success) ~ 0.47) was much lower than the rate presented by [MacGregor \(2017\)](#) (0.66 to 0.91), and our MCMC Binomial Bayesian analysis, using the information provided by MacGregor as a prior, showed a likelihood estimate of the participant selecting the most optimum solution as being 0.54 (95% credible interval [0.46, 0.63]).

We mined geometric properties from the ETSP solutions generated for our experiment and used them in two models (a mixed-effects model, and a drift-diffusion model) to both evaluate the importance of the variables in predicting human discrimination of optimality between competing tours, and to assess to what extent measures of area, complexity, regularity, and tour costs can be used as predictors of stimulus selection. The study looked at the effect of measures when predicting optimality amongst competing stimuli. Interestingly, although our models can accurately predict the most optimal stimulus in trials 96% of the time (on previously unseen data), the same geometric properties only explain a very small amount of the variance in participant selections ($\sim 14\%$), with another 6% being down to random effects. Ultimately there is definitely more ‘going on’ here, which we aim to investigate in future work. To this end, we have made our data and our scripts available online for the sake of replicability, as well as to generate additional interest in the area (see [Appendix 2](#) for the URL).

The results from the diffusion-drift models was promising, and we certainly plan to use this method more in future studies, however much work is needed to fully identify the factors impacting human selection between pre-existing ETSP solutions. We aim to undertake future research that uses the fact that all geometric properties correlate, to

explore the effects on human selection in a more controlled experiment. Moreover, consideration of personal, information assimilation, and individual difference factors are planned in future work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix 1

A.1. Approximating area using the Monte Carlo method

Since we are using arbitrary units of measurement, we took the ratio of the area between the complex polygons created by the graph solutions, and the total area of the graph, as an appropriate measure of area for the TSP solutions. To get a good approximation of the area of each polygon, we used the Monte Carlo method to generate a large number of points that were randomly placed on the graph, and then took the proportion of the points that fell within the boundaries of the polygon over the total number of points. Descriptively, the way this method works is by projecting horizontal lines from each point and using a system of equations to find all intersections with the projected lines of the edges. If an intersection is found, the algorithm then checks to ensure that the intersection is within the boundaries of the edge. Finally, if the number of total intersections is an even number, then the point is outside the polygon, if it is odd, then the point is within the polygon.

More formally, suppose a set of points $P = (X, Y)$ is generated such that $X = X_1, \dots, X_n \sim U(a, b)$ and $Y = Y_1, \dots, Y_n \sim U(a, b)$ where $U(a, b)$ denotes a uniform distribution with a minimum number a and a maximum number b (which in our experiment ranged from 0 to 100). Furthermore, for the sake of notational clarity, x and y were defined as mapping functions i.e., $x, y : \mathbb{R}^2 \rightarrow \mathbb{R}$, so that they decompose the two-dimensional point to its x and y coordinates. For example, for the point $T = (52, 45)$, $x(T)$ is 52, and $y(T)$ is 45. For illustration purposes, only one such point is presented (for now) in Fig. A1.

To identify if point P has fallen within the polygon, initially a horizontal line is ‘drawn’ from the point P . The linear form of the line is

$$y_I = 0 + y(P), \quad (\text{vi})$$

where y_I is the y coordinate of the point of intersection between the two lines (see Fig. A2).

Next, the linear projections of all TSP solution edges are extracted. Each pair of points, which make up the edge, are defined as being $H_i = (G_i, G_{i+1})$. Extraction of edges is an iterative process, however for the sake of clarity we consider a single pair of points (i.e., H_1). To generate the line for each pair of points, the slope is determined

$$m = \frac{y(G_2) - y(G_1)}{x(G_2) - x(G_1)} | x(G_2) \neq x(G_1). \quad (\text{vii})$$

Identification of the slope allows the intercept to be determined

$$c = y(G_1) - m^*x(G_1). \quad (\text{viii})$$

Finally, we combined the two equations, since the value of y_I is already known, allowing us to find x_I . By rearranging the formula $y = mx + c$, i.e.:

$$x_I = \frac{y - c}{m} | m \neq 0, \quad (\text{ix})$$

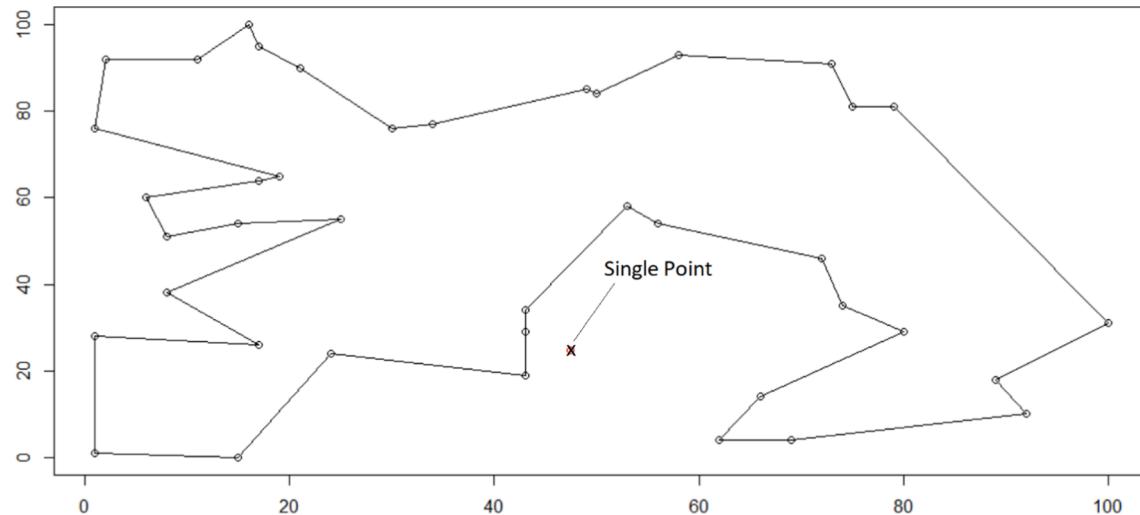


Fig. A1. A randomly placed point $P = (x, y)$ in the graph, with coordinates x and y sampled from a uniform distribution with $\min = 0$ and $\max = 100$.

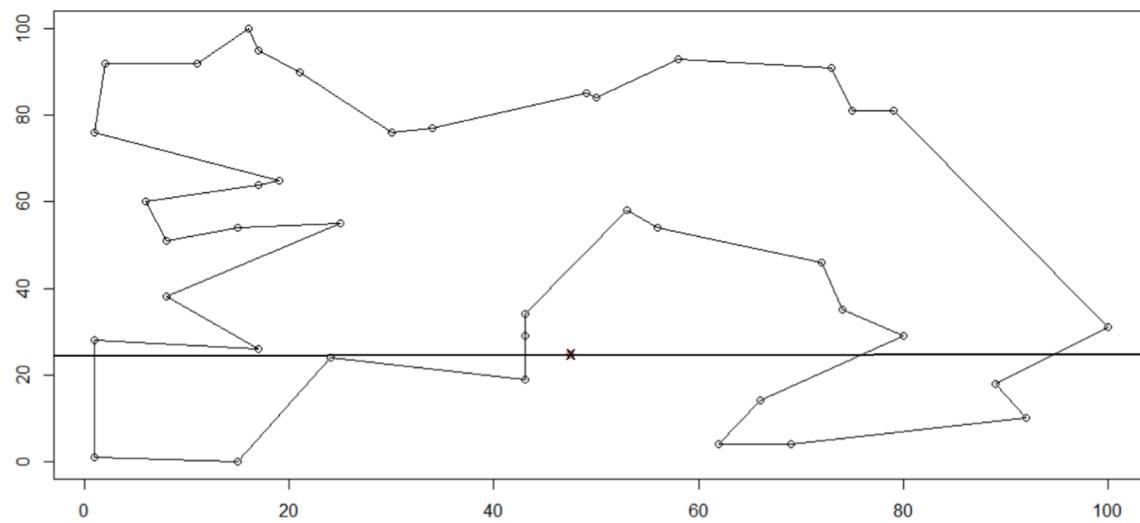


Fig. A2. A horizontal line ‘drawn’ on the point P , so that $y_I = y(P)$.

the intersection point (x_I, y_I) is determined. This works well in most cases, however, there are two exceptions to this rule. The first condition, as shown in (vii), is if the edge projects a vertical line, then there will be no slope, i.e., $x(G_2) = x(G_1)$. This happens more regularly when $G \in \mathbb{N}^2$, as was the case with our graphs. In these cases, x_I is set as

$$x_I = x(G_2), \quad (x)$$

and y_I as

$$y_I = y(P). \quad (xi)$$

The second condition is that the slope of the edge cannot be equal to 0, shown in (ix), i.e., the edge cannot be horizontal. If that is the case, then the result would be undefined, since you would have parallel lines that would never intersect. Algorithmically, one solution is to return a non-meaningful coordinate, e.g., containing negative numbers, which is outside the bounds of our problem space.

Implementation of the above algorithm will return all intersections with the horizontal line, however not all intersections are meaningful, since we are only interested in finding the intersections within the bounds of the edges themselves (not their linear projections). To check if an intersection is within the bound of the TSP problem space, an indicator function can be used. Suppose the intersection point $I = (x_I, y_I)$, then,

$$g(I, G) = \begin{cases} 1 & \text{if } x(G_1) < x(I) < x(G_2) \text{ and } y(G_1) < y(I) < y(G_2) \\ 0 & \text{Otherwise} \end{cases}. \quad (xii)$$

Algorithmically, optimisation of the process can be achieved determining whether a graph point is within the polygon by checking intersections to either the left ($x(P) < b/2$) or the right ($x(P) \geq b/2$); where b is the maximum number from the uniform distribution of points that were used to generate the graph (i.e., 100). By extending this iteratively to all edges in the graph the result can be determined (see Fig. A3).

If the number of intersections is even, then the point must lay outside the polygon. If the number of intersections is odd, the point is inside the polygon. Formally, assigning that number to a variable

$$v = \sum g(I) \quad (xiii)$$

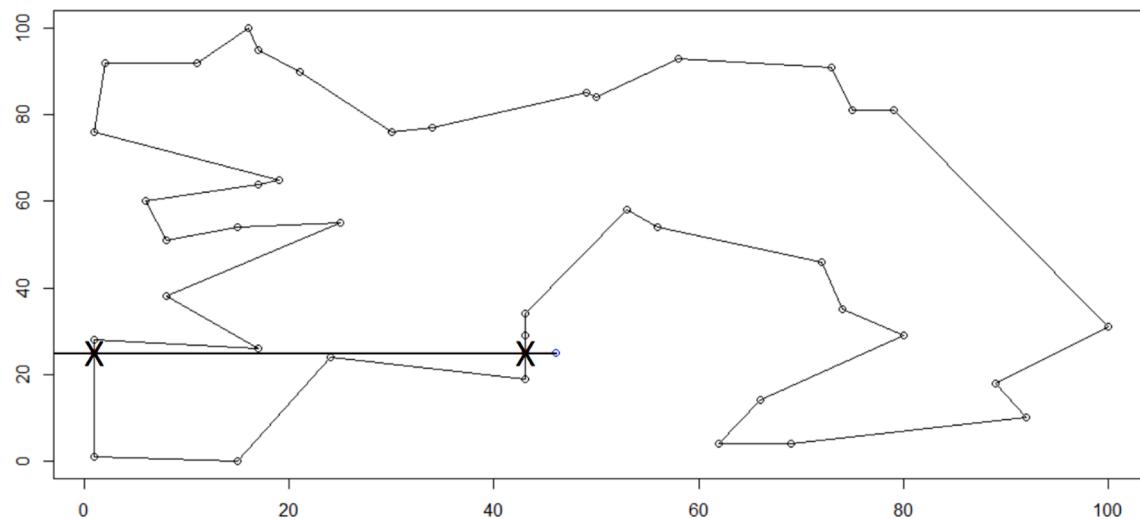


Fig. A3. Process is optimised by only counting the number of intersections to the left of the point if $x(P) < a/2$ and to the right if $x(P) \geq a/2$.

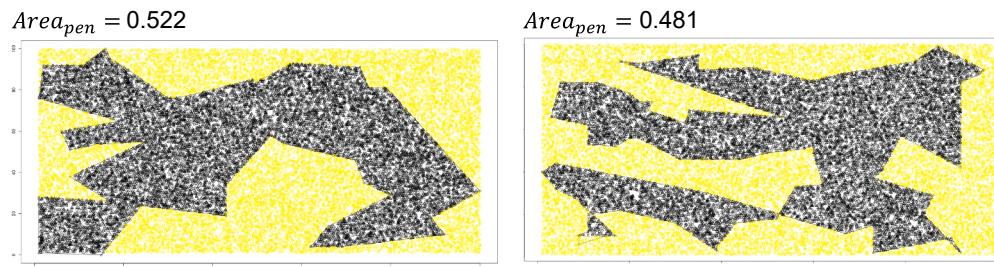


Fig. A4. Graph containing no crossings (left), and crossings (right). The area is extracted using a Monte Carlo method and is simply the ratio of points inside the polygon to the total number of points.

Table A1

Examples of differences in $Area_{pen}$ estimates derived by the function $w(s, n)$.

Graph ID	Number of Nodes (n)		
	n = 10,000	n = 20,000	n = 30,000
5	0.51	0.51025	0.5080667
27	0.37	0.364	0.3646333
150	0.5534	0.55775	0.5605333
300	0.5621	0.5487	0.5522667

Note. The Graph ID identifies solutions in our dataset, which are made available on Github (see [Appendix 2](#), for the URL). Also, the ratio measurements of area are point estimates, and can differ slightly due to the stochastic nature of Monte Carlo. The algorithm works well for graphs with, and without crossings in most cases, as shown in [Fig. A4](#).

With our final indicator function, the modulus of v is returned, which will inform us whether the point is inside or outside the polygon, i.e.

$$o(v) = \begin{cases} 1 & \text{if } v \bmod 2 = 1 \\ 0 & \text{otherwise} \end{cases}. \quad (\text{xiv})$$

The procedure can be used with an arbitrarily large number of points, allowing us to determine the number of points that fall within the polygon using the function $o(v)$. In other words, if we were to create n number of points, and run process steps (vi) to (xiv) n number of times, each time updating $x(P)$ and $y(P)$, then the sum of the $o(v)$ outputs can be used to assign a value representing number of points inside the polygon,

$$\tau = \sum_{i=1}^n o(v). \quad (\text{xv})$$

As we increase n , results tend towards a better estimate of the area of the polygon by taking the ratio

$$w(r, n) = \frac{\tau}{n} = Area_{pen}, \quad (\text{xvi})$$

where w is a function of τ and n , and returns an estimate of the true area of the polygon, which can be denoted as θ . Naturally, $\lim_{n \rightarrow \infty} w(\tau, n) = \theta$, however, this iterative process (i.e., our Monte Carlo method), is computationally expensive. For one forty-node graph, it took the University Academic Cluster approximately four seconds to compute w when $n = 5000$, eight seconds when $n = 10,000$, and 15 s when $n = 20,000$. Interestingly, there is a point of saturation, i.e., a compromise between accuracy and speed. We suggest that a good value for n is 20,000, which is approximately when the difference in output precision is equivalent at two decimal places (see [Table A1](#)).

Finally, as discussed earlier, we suggest that areas for solutions that contain internal crossings should be penalized. Our approach penalizes these graphs by not considering the internal area created by crossings (as shown in [Fig. 5](#) of the main text).

A.2. Calculating the number of edge crossings in solutions

The graphs are made up of nodes, which when connected form edges. These edges are straight lines and can be projected to follow the usual linear equation $y = mx + c$ by getting the slope and the intercept as shown in steps (vii) and (viii). To check if two edges in the graph intersect, one would need to (a) find where the intersection is between the two lines, followed by (b), check if that intersection falls within the boundary space of the edges. Finding the intersections is another iterative process that requires the checking of every edge against every other edge. This can be expressed using two functions. The first function checks where the linear projections of the lines intersect using a system of two equations

$$f_c(l_1, l_2) = \begin{cases} y = ax + c \\ y = bx + k \end{cases}. \quad (\text{xvii})$$

In the case of two vertical lines, two horizontal lines, or horizontal and vertical lines, we need to make exceptions. In the first two cases, our function would be undefined, so algorithmically coordinates that are outside the problem space can be returned as a negative number. In the last case the steps shown in (x) and (xi) would be followed.

As an example, let us consider

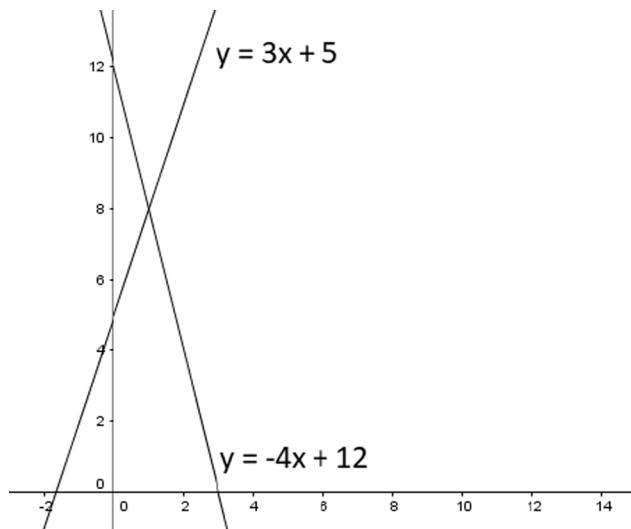


Fig. A5. Two linear projections of edges will cross unless they are parallel lines.

$$\begin{cases} y = 3x + 5 \\ y = -4x + 12 \end{cases} \quad (\text{xviii})$$

The intersection point is shown in [Fig. A5](#).

This expressed in standard form allows us to align x and y

$$\begin{cases} -3x + y = 5 \\ 4x + y = 12 \end{cases} \quad (\text{xix})$$

Next, multiplying the equation on the bottom by $\frac{3}{4}$ in order to get $x = 3$, would allow the removal of x when adding the two equations.

$$\begin{cases} -3x + y = 5 \\ \frac{12x}{4} + \frac{3y}{4} = \frac{36}{4} \end{cases} \quad \begin{cases} -3x + y = 5 \\ 3x + \frac{3y}{4} = 9 \end{cases} \quad (\text{xx})$$

This allows us to solve y

$$(-3x + 3x) + \left(\frac{3y}{4} + y\right) = (9 + 5) \quad (\text{xxi})$$

$$0 + \frac{7y}{4} = 14$$

$$\frac{4}{7} \left(\frac{7y}{4}\right) = \frac{4}{7} 14$$

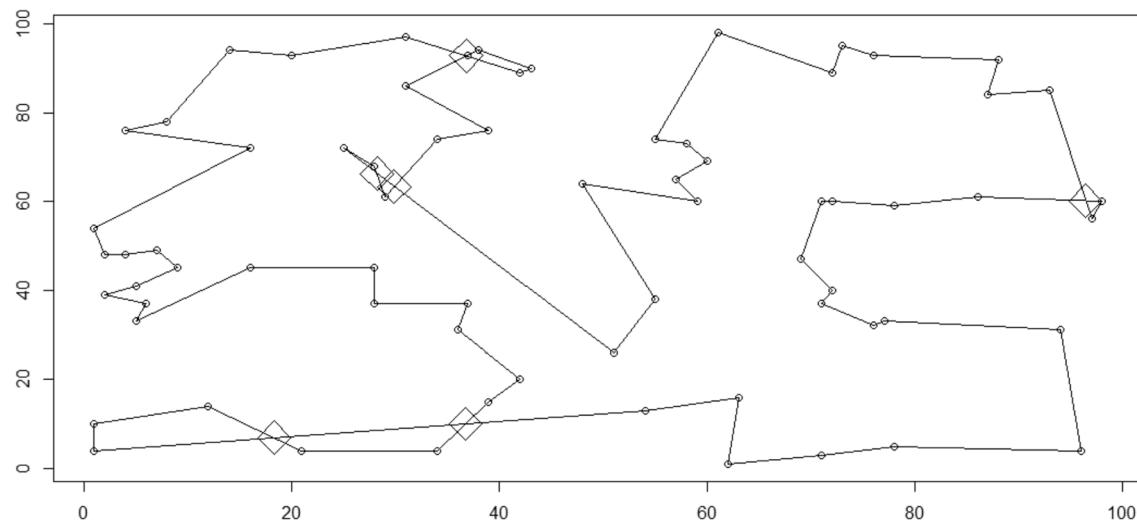


Fig. A6. Our algorithm detected the six intersections in this graph using an iterative system of two linear equations.

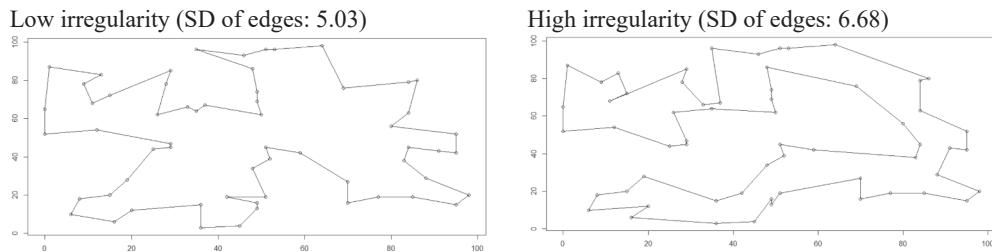


Fig. A7. According to our analysis, the graph on the left is more regular than the one on the right (edges are more likely to be similar lengths).

$$y = 8,$$

which can be replaced in either of the two equations in order to identify x

$$8 = 3x + 5 \quad (\text{xxii})$$

$$3x = 8 - 5$$

$$x = (8 - 5)/3$$

$$x = 1.$$

Therefore, the intersection point in this example is $(1, 8)$. By iterating through this process, we were able to identify all intersections and add them to the set of intersection points I_{int} .

The second part of the process is to check whether the intersection points are within the bounds of both edges. This is similar to process step (xii), however one would need to run the function twice – once for each edge. Let us imagine two edges $(G^a$ and G^b) exist. Each edge is made up of two nodes, i.e., $G_1^a = (x_1, y_1)$ and $G_2^a = (x_2, y_2)$. The function shown in (xii) is required to check if the intersection is within the bounds of the first edge $g(I_{int}, G^a)$, and then to check if it is within the bounds of the second edge $g(I_{int}, G^b)$. Again, a simple indicator function can be used, which will sum the in-bound intersections later

$$q(g(I_{int}, G^a), g(I_{int}, G^b)) = \begin{cases} 1 & \text{if } vg(I_{int}, G^a) = 1 \text{ and } g(I_{int}, G^b) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (\text{xxiii})$$

Finally, intersection points can be counted by considering the number returned by function q .

$$n_{inter} = \sum_{i=1}^n q(g(I_{int}, G^a), g(I_{int}, G^b)). \quad (\text{xxiv})$$

For illustration purposes, we demonstrate the proposed algorithm on a graph solution containing six crossing; highlighted by marked line intersections (see [Fig. A6](#)).

Interestingly, despite the simplistic nature of the problem, we could not find a script specifically designed to either i) detect intersections or ii) measure areas in the TSP. Therefore, two R scripts were created to facilitate the process and have been shared on Github (see [appendix 2](#) for the URL).

A.3. Cost of TSP tours

The total cost of a tour is the sum of all distances travelled between nodes in the solution (i.e., the edges). Since every edge is made up of two nodes, which we denote as $G_i = (x_i, y_i)$ and $G_{i+1} = (x_{i+1}, y_{i+1})$, the Euclidean distance d of the tour is

$$d_{tot} = \sum_{i=1}^{n-1} \sqrt{(x(G_i) - x(G_{i+1}))^2 + (y(G_i) - y(G_{i+1}))^2}. \quad (\text{xxv})$$

A.4. Regularity of TSP solutions

A polygon is regular when all sides and all angles are equal. An irregular polygon will have one or more incongruent sides. TSP solutions can therefore be seen as complex irregular polygons. The variability in the length of edges can be extracted by taking the standard deviation. As a reminder, every graph G , is made up of an arbitrary number of nodes n where $\{n : n \in \mathbb{N}, n > 2\}$, denoted as G_1, \dots, G_n , and every edge H_i is made up of the two adjacent points (G_i, G_{i+1}) . To get the variability of edge lengths, the standard deviation can be determined using

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} \left(\left(\sqrt{(x(G_i) - x(G_{i+1}))^2 + (y(G_i) - y(G_{i+1}))^2} \right) - \frac{1}{n} d_{tot} \right)^2 \quad (\text{xxvi})$$

$$s = \sqrt{s^2}.$$

It is expected that each edge is likely to vary in length when compared to other edges in the solution, however we hypothesise that as variation increases, the solution is more irregular (see [Fig. A7](#)).

Appendix 2

Scripts and datasets for this study are available at: https://github.com/markoskyritsis/TSP_Scripts.

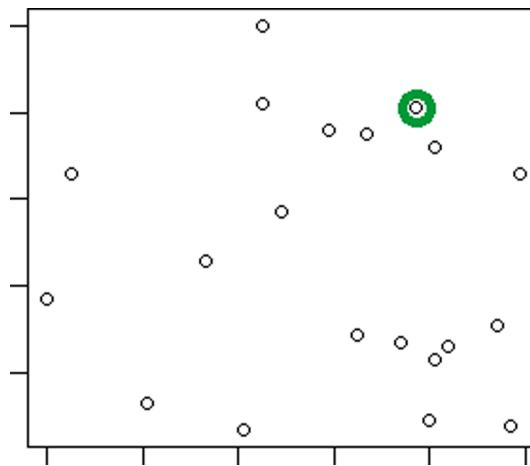
Appendix 3

Below are the instructions given to participants (converted from HTML for presentation purposes) prior to the start of the experiment (just after clicking 'consent').

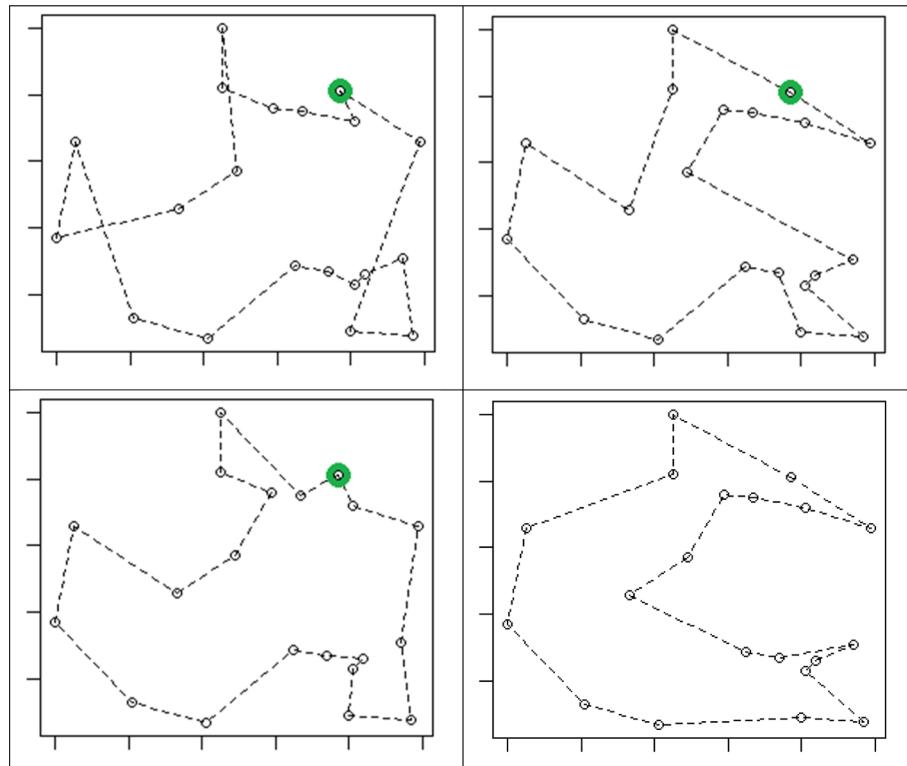
Page1

A short tutorial

The travelling salesperson problem describes a hypothetical scenario where a salesperson wants to travel from their home city to multiple other locations once and only once before returning home. There are multiple tours the salesperson could take, some more optimal than others. Consider the following graph (starting city circled).



Below are four possible tours the salesperson could have taken. Some more optimal than the others

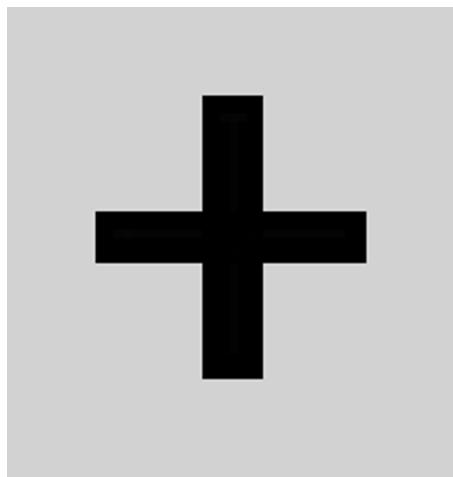


In this case, our analysis shows that the tour on the bottom-left is the most optimal in this set, since it requires less travelling. We are interested in how good you are at 'guessing' which tour is most optimal just by looking at four possible solutions.

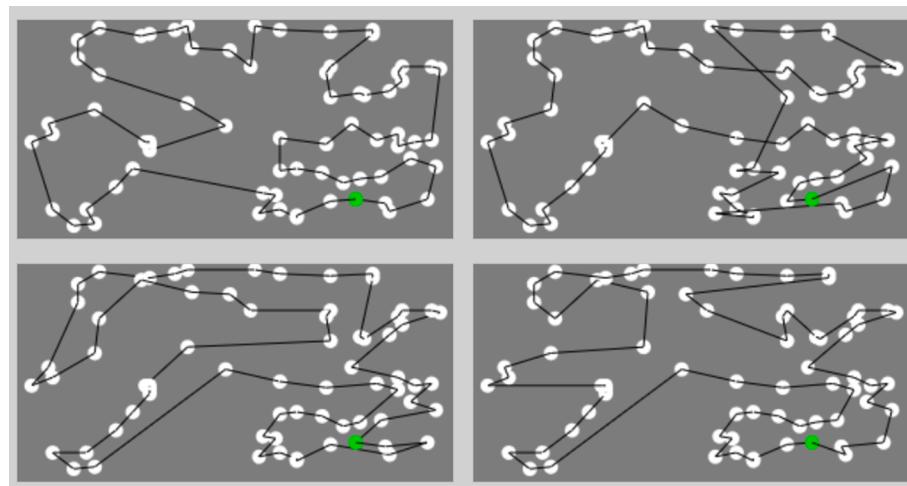
Page 2

The process

The experiment consists of 50 trials. Each trial will start with a crosshair



After a second or so, you will be presented with four tours for a random graph. It's up to you to decide which of those you think is the best tour. In other words, is the tour that the salesperson travelled the least distance



Each trial is timed, so try your best to complete it is as quickly as possible, without compromising the quality of your decision.

References

Alexandrowicz, R. W. (2020). The diffusion model visualizer: An interactive tool to understand the diffusion model parameters. *Psychological Research Psychologische Forschung*, 84(4), 1157–1165.

Barton, K. (2020). MuMin: multi-model inference. <http://r-forge.r-project.org/projects/mumin/>.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Blaser, R. E., & Wilber, J. (2013). A comparison of human performance in figural and navigational versions of the traveling salesman problem. *Psychological Research Psychologische Forschung*, 77(6), 761–772.

Dry, M., Lee, M. D., Vickers, D., & Hughes, P. (2006). Human performance on visually presented traveling salesperson problems with varying numbers of nodes. *The Journal of Problem Solving*, 1(1), 4.

Dry, Matthew, Preiss, Kym, & Wagemans, Johan (2012). Clustering, randomness and regularity: Spatial distributions and human performance on the traveling salesperson problem and minimum spanning tree problem. *The Journal of Problem Solving*, 4(1). <https://doi.org/10.7771/1932-6246.1117>

Dry, M. J., & Fontaine, E. L. (2014). Fast and Efficient Discrimination of Traveling Salesperson Problem Stimulus Difficulty. *The Journal of Problem Solving*, 7(1), 9.

Friedenberg, J., & Bertamini, M. (2015). Aesthetic preference for polygon shape. *Empirical Studies of the Arts*, 33(2), 144–160.

Graham, S. M., Joshi, A., & Pizlo, Z. (2000). The traveling salesman problem: A hierarchical model. *Memory & Cognition*, 28(7), 1191–1204.

Hahsler, M., & Hornik, K. (2007). TSP-Infrastructure for the traveling salesperson problem. *Journal of Statistical Software*, 23(2), 1–21.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... & Team, R. C. (2018). caret: Classification and Regression Training. R package version v6. 0.77. 2017.

Kyritsis, M., Gulliver, S. R., & Feredoes, E. (2018a). Acknowledging crossing-avoidance heuristic violations when solving the Euclidean travelling salesperson problem. *Psychological Research Psychologische Forschung*, 82(5), 997–1009.

Kyritsis, M., Gulliver, S. R., Feredoes, E., & Din, S. U. (2018b). Human behaviour in the Euclidean Travelling Salesperson Problem: Computational modelling of heuristics and figural effects. *Cognitive Systems Research*, 52, 387–399.

MacGregor, J. N., & Chu, Y. (2011). Human performance on the traveling salesman and related problems: A review. *The Journal of Problem Solving*, 3(2), 2.

MacGregor, J. N., & Ormerod, T. (1996). Human performance on the traveling salesman problem. *Perception & Psychophysics*, 58(4), 527–539.

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(2), 260–270.

MacGregor, J. N. (2012). Indentations and starting points in traveling sales tour problems: Implications for theory. *The Journal of Problem Solving*, 5(1), 3.

MacGregor, J. N. (2017). Heuristics for Comparing the Lengths of Completed E-TSP. Tours: Crossings and Areas.

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9, 9, 2014.

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.

Pizlo, Z., Stefanov, E., Saalwechter, J., Li, Z., Haxhimusa, Y., & Kropatsch, W. G. (2006). Traveling salesman problem: A foveating pyramid model. *The Journal of Problem Solving*, 1(1), 8.

R Core Team, (2016). Vienna: R Foundation for Statistical Computing, 2016. R: A language and environment for statistical computing.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333.

Smith, P. L., & Ratcliff, R. (2015). *An introduction to the diffusion model of decision making. In An introduction to model-based cognitive neuroscience* (pp. 49–70). New York, NY: Springer.

Van Rooij, I., Stege, U., & Schactman, A. (2003). Convex hull and tour crossings in the Euclidean traveling salesperson problem: Implications for human performance studies. *Memory & Cognition*, 31(2), 215–220.

van den Bergh, Tuerlinckx, & Verdonck. (2020). DstarM: an R package for analyzing two-choice reaction time data with the D* M method. *Behavior Research Methods*, 52(2), 521–543. <https://doi.org/10.3758/s13428-019-01249-7>

van Rooij, I., Schactman, A., Kadlec, H., & Stege, U. (2006). Perceptual or Analytical processing? Evidence from children's and adult's performance on the Euclidean traveling salesperson problem. *The Journal of Problem Solving*, 1(1), 6.

Verdonck, S., & Tuerlinckx, F. (2016). Factoring out nondecision time in choice reaction time data: Theory and implications. *Psychological Review*, 123(2), 208–218. <https://doi.org/10.1037/rev0000019>

Vickers, Lee, Dry, & Hughes. (2003). The roles of the convex hull and the number of potential intersections in performance on visually presented traveling salesperson problems. *Memory & Cognition*, 31(7), 1094–1104.

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

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, 32(7), 1206–1220.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of values. *Psychonomic Bulletin & Review*, 14(5), 779–804.

Wagenmakers, E. J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, 21(5), 641–671.