# Cognitive Decision-Making in TSP Tasks: The Impact of Line Stylization Features of Point Arrays

**Chen Chen (6223112008@stu.jiangnan.edu.cn)**
School of Artificial Intelligence and Computer Science,Jiangnan University, 1800 Lihu Avenue, Binhu District
Wuxi, Jiangsu 214122, China

**Ruimin Lyu* (Corresponding author, ruiminlyu@jiangnan.edu.cn )**
School of Artificial Intelligence and Computer Science,Jiangnan University, 1800 Lihu Avenue, Binhu District
Wuxi, Jiangsu 214122, China

## Abstract

The Traveling Salesman Problem (TSP) is a classic NP-hard problem, and research on its cognitive decision-making often focuses on internal factors like memory and experience, while neglecting the influence of the problem's structural characteristics. This study identifies that potential linear features in the TSP point distribution (such as implied paths formed by visual aggregation) may significantly impact human path selection strategies. To test this hypothesis, we propose a method for quantifying the Line Stylization Degree and generate different TSP instances with varying characteristics by introducing disturbances. These are then combined with experimental analysis of participants' decision-making patterns. The results show that participants tend to plan paths along implied lines, and this strategy may reduce cognitive load. The contribution of this paper lies in revealing the shaping role of visual structural features on cognitive decision-making, providing theoretical support for designing human-centered path planning algorithms.

**Keywords:** TSP,Line stylization features,Cognitive Decision Making,Human Performance

## Introduction

The Traveling Salesman Problem (TSP)(Flood, 1956), a classical NP-hard problem in computer science and operations research, holds significant theoretical importance and serves as a canonical example in path planning studies. The inherent complexity of TSP has established it as an ideal benchmark for testing and optimizing diverse algorithms, while its broad applicability extends to real-world domains such as logistics, transportation, and network design. Consequently, research on TSP not only carries profound academic implications but also demonstrates substantial practical value.

In cognitive science research on the Traveling Salesman Problem (TSP), numerous scholars have investigated the strategic choices humans employ when solving it, particularly the roles of cognitive factors such as memory, attention, and experience. For instance, studies indicate that humans often rely on heuristic strategies to simplify the problem-solving process, applying rules to reduce computational complexity(Bickley & Torgler, 2021). However, existing research has predominantly focused on behavioral patterns, strategy selection during problem-solving, and how these strategies are influenced by intrinsic cognitive factors(MacGregor & Ormerod, 1996; Graham et al., 2000; Vickers et al., 2003; Dry et al., 2006; MacGregor et al., 2006). Comparatively few studies have explored how the structural features of TSP instances themselves(MacGregor et al., 1999)—especially the geometric properties of spatial node distributions—may implicitly shape human decision-making processes.

We have observed that in TSP instances with a large number of cities, the spatial distribution of nodes may form latent line stylization features through visual aggregation effects. Specifically, certain discrete point arrays may be perceptually organized into implied linear structures—patterns not explicitly defined as paths but emerging from aggregated visual perception. For example, when confronted with a problem configuration as illustrated in Figure 1, humans might adopt the path-crossing avoidance strategy previously identified in TSP studies, potentially favoring the connection scheme shown in Solution A to minimize path intersections(Rooij et al., 2003). However, we note that the problem configuration inherently contains latent line stylization features. If paths are connected according to these linear patterns, intersections would inevitably arise. This raises the question: which scheme—adherence to line stylization features or strict avoidance of crossings—would yield higher efficiency in such scenarios?

This phenomenon raises two critical questions: First, when the spatial distribution of nodes in a TSP instance exhibits line stylization features, do participants exhibit a tendency to align their route planning with these implied linear patterns? Second, does adherence to such latent linear structures during path selection lead to improved task efficiency or reduce cognitive load for participants when solving the problem? To address these questions, we designed and conducted

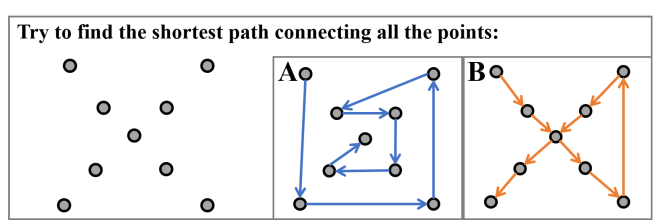


Figure 1: When faced with the TSP problem in the diagram, would you choose option A or option B?

a series of experiments. First, we quantified the degree of line stylization in the spatial distribution of nodes within TSP configurations. By introducing controlled perturbations of varying magnitudes, we generated TSP instances with sys-



In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* 

tematically modulated line stylization features. These perturbations enabled precise control over the salience of line stylization characteristics, thereby allowing systematic evaluation of their impact on path selection. Subsequently, we recruited multiple participants to solve TSP problems under these distinct configurations and performed detailed analyses of their decision-making processes and path selection strategies. Specifically, we examined whether participants exhibited a tendency to align their routes with perceived linear patterns during problem-solving and whether such strategies enhanced their task efficiency or solution accuracy.

## Method

To investigate our research questions, we designed and implemented a series of TSP problem configurations and applied perturbations of varying magnitudes to generate control groups with significantly distinct degrees of line stylization. By manipulating the geometric properties of point distributions, these configurations exhibited varying levels of visual line stylization features. We designed two experiments: Experiment 1 aimed to quantitatively measure the changes in the ”degree of line stylization” of the designed TSP instances before and after perturbation. We developed a novel method to perform quantitative analysis of point arrays across configurations, deriving specific numerical values for the degree of line stylization. Experiment 2 was designed to measure human performance levels on the constructed TSP problem set. Based on the TSP-solving process, we developed an experimental procedure to test participants’ performance and efficiency in solving TSP problems under different configurations.This experiments were conducted with prior approval from the institutional ethics review board (Grant No. JNU202409RB0028).

### TSP Set Design

In this experiment, a total of 30 TSP problem configurations were used (see Figure 2), including one set of original configurations and two sets of configurations perturbed to varying degrees. Each set contained 10 problem instances. The original configurations were generated by discretizing symbols and lines of varying complexity to ensure moderate structural complexity and visual discernibility. The perturbed configurations introduced distinct line stylization features by either enhancing or reducing the interconnectivity between points.

To investigate the influence of the degree of line stylization in point arrays on TSP problem-solving, we designed two categories of perturbations with distinct magnitudes to generate experimental configurations exhibiting different line stylization features. By introducing perturbations, we systematically altered the structural properties of point arrays while preserving the overall TSP problem configuration, thereby enabling controlled manipulation of the degree of line stylization. Specific perturbation methods included random point deletion and random positional shifting of points, As shown in Figure 3.both of which induced varying degrees of line stylization feature alterations.

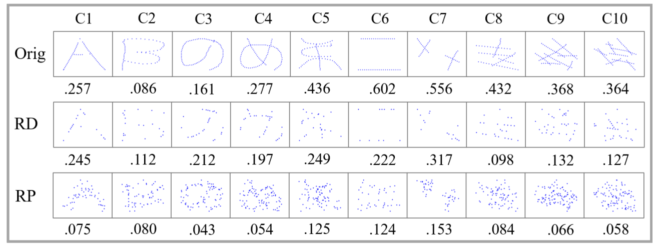


Figure 2: The specific distribution of the 30 TSP problem lattice configurations used in the experiment, where each column represents three configurations for a single problem, numbered from C1 to C10.”Orig” represents the Original group, ”RD” represents the RandomDelete group, and ”RP” represents the RandomPosition group.The Fleiss’ Kappa value for this configuration is annotated at the bottom of the figure.

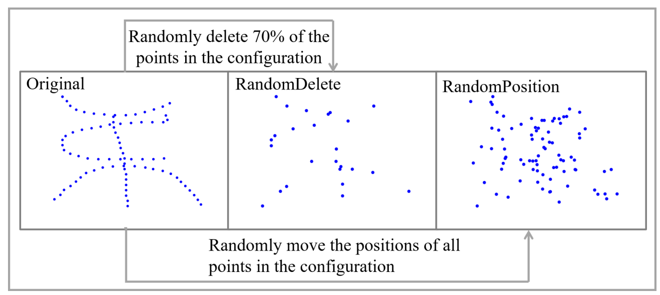


Figure 3: An example of the perturbation of the original TSP problem configuration.

### Participants

Experiment 1 involved 130 undergraduate students, while Experiment 2 had 17 participants, all of whom were adults over the age of 18. None of the participants had been exposed to the TSP problem configurations used in the experiments prior to the start, and all possessed a certain level of computer proficiency. Before the experiments began, we provided detailed operational instructions to the participants, ensuring they fully understood the task requirements and process, and were able to use the experimental program proficiently.

### Experiment 1: Quantification of Line Stylization Degree in TSP Task Sets

**Experimental Objective and Design** The objective of this experiment was to measure the degree of line stylization in point arrays within each TSP problem configuration. To operationalize this measurement, we first defined it as a subjective quantity and adopted the following core principle: A set of points is considered to exhibit a high degree of line stylization if multiple human observers consistently perceive and connect the points into linear structures, with significant agreement in their connection patterns.

Guided by the aforementioned principle, we instructed participants to annotate ”lines” within the TSP problem set proposed in this study. Each participant connected clusters of points they perceived as belonging to the same linear structure through line-drawing based on their cognitive judgments. By analyzing the consistency of participants' line categorizations across each problem configuration, we quantified the degree of line stylization for each configuration and assigned a specific line stylization score. This quantification method serves as the foundation for subsequent decision-behavior analyses and facilitates understanding of how distinct line stylization features may exert latent influences on path-planning strategies.

**Procedure** The experimental interface is illustrated in Figure 4. In each experimental stage, after entering personal information, participants were presented with randomly ordered point array configurations of varying line stylization features. Each participant completed 26 annotation tasks. Detailed operational instructions were displayed on the right side of the interface to guide participants through task procedures. This approach enabled participants to externalize their visual and cognitive judgments, thereby generating data to quantify the degree of line stylization in the point arrays.

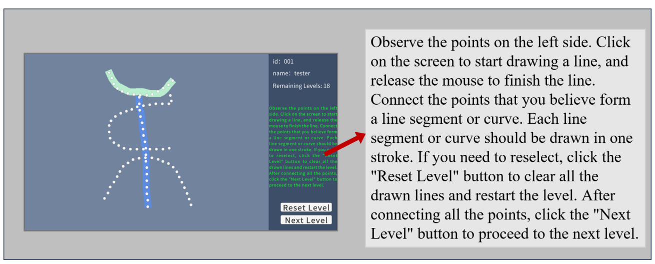


Figure 4: The experimental program interface and operation instructions for Experiment 1.

## Experiment 2: Measuring Human Performance on the TSP Task Sets

**Experimental Objective and Design** The objective of this experiment was to investigate human performance in TSP tasks under configurations with varying degrees of line stylization.

**Procedure** The core task of this experiment is to have participants draw paths using a computer program, connecting all the target points in the TSP problem configuration (as shown in Figure 5). After entering their personal information, each participant is randomly presented with a TSP problem configuration and is asked to connect all the target points according to what they believe to be the shortest path. To ensure that participants can fully explore and optimize their path planning, they are allowed to undo previously drawn paths during the process, making it easier to adjust their decisions.

Unlike traditional TSP experiments, which typically employ mouse-clicking to connect target points, our experimental design adopted a manual path-drawing approach. This decision was grounded in the hypothesis that freehand drawing and writing—common activities in daily experience—align more closely with cognitive mechanisms associated with linear pattern perception, thereby enhancing the ecological validity of the task in studying how line stylization features influence decision-making(Mueller & Oppenheimer, 2014).

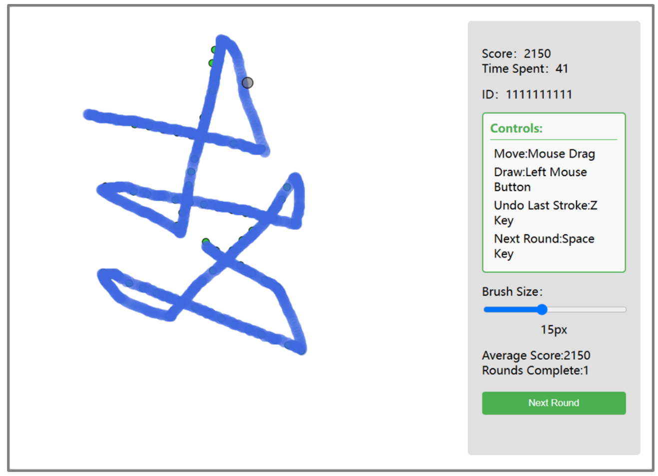


Figure 5: The experimental program interface for Experiment 2.

# Results and Analysis

## Experimental Data Description

In Experiments 1 and 2, the number of human annotations and solutions obtained for each TSP task was M=44 (SD=±7) and M=17 (SD=±0), respectively. Each dataset in Experiment 1 included participants' line annotation data for each TSP task. Each dataset in Experiment 2 documented participants' path-planning processes, path selections, and time spent to resolve the problem across all 30 problem configurations.

## Calculation of Line Stylization Degree

This section aims to calculate the degree of line stylization for each TSP problem configuration in the proposed dataset. In Experiment 1, multiple human participants performed line categorization across different TSP configurations. By quantifying the consistency of participants' line categorizations for each configuration, we derived an approximate degree of line stylization for each problem instance. To formalize this quantification, Fleiss' Kappa—a statistical method widely used to assess inter-rater agreement when multiple evaluators classify categorical items—was employed(Fleiss, 1971; Fleiss et al., 1981). Fleiss' Kappa extends Cohen's Kappa(Cohen, 1960) to scenarios involving more than two evaluators, measuring the level of agreement among participants relative to chance. Specifically, it quantifies the discrepancy between observed agreement and random agreement, thereby reflecting the consensus among participants in categorization tasks.Est´evez-Almenzar et al. employed the Fleiss' Kappa method to assess the level of agreement among multiple human raters in facial recognition tasks(Estévez-Almenzar et al., 2025). Similarly, (Goh et al. utilized Fleiss'

Kappa in their study to evaluate the consistency between human and machine performance in text classification tasks. These examples demonstrate that the Fleiss' Kappa method is well-suited for assessing consistency in human classification tasks(Goh et al., 2020). Additionally, Huang et al. illustrated the constructive role of "actional involvement" in line perception by analyzing the dependence, variability, and dynamic feedback of brushstroke operations in line control. By applying Fleiss' Kappa, we rigorously measured the consistency of participants' line categorizations across all TSP configurations, generating a numerical line stylization score for subsequent analyses.First, for each TSP configuration, the frequency of classification of points within the same line by all participants is statistically analyzed to calculate their observation consistency. Then, the average observation consistency across all configurations is computed to evaluate the overall classification stability. Next, the expected consistency is calculated based on the probability of random classification, representing the baseline level of categorization without any pattern. Finally, Fleiss' Kappa formula is used to compare observation consistency with expected consistency, quantifying the degree of inter-rater agreement.The formula is as follows:

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \quad (1)$$

Where $\bar{P}$ represents the observed consistency and $\bar{P}_e$ represents the expected consistency. Using this formula, we derived a line stylization score for each TSP problem configuration. These scores quantify the structural characteristics of the configurations and assess their association with decision-making behavior.

Figure 2 presents the line stylization scores for the 30 TSP problem configurations in Experiment 2. These scores reflect the degree of line stylization in point array configurations under different perturbation conditions and provide a foundational basis for subsequent analyses of path-planning efficiency.

## Impact of Line Stylization Features on Human Performance

This section aims to explore the impact of line stylization features on human performance in solving the Traveling Salesman Problem (TSP). To rigorously evaluate the quality of human-generated solutions, we assessed each solution based on two metrics: the PAO value (Percentage Above Optimal) of path length (i.e., the percentage deviation of the path length relative to the optimal path length) and the number of path intersections. The PAO value quantifies the excess path length compared to the optimal solution(MacGregor & Chu, 2011; MacGregor, 2015; Rach & Kirsch, 2016), with its core purpose being to measure the proximity of the current solution to the theoretical optimum. This metric is commonly used to evaluate solution efficiency. Since our TSP instances do not guarantee optimal solutions, the PAO value serves as a standardized measure to normalize performance across different problem configurations, enabling a unified assessment of human participants' performance. The Lin-Kernighan-Helsgaun (LKH) algorithm, one of the most powerful and efficient algorithms for solving TSP(Helsgaun, 2000; Zheng et al., 2023; Kerschke et al., 2018), consistently produces near-optimal solutions in most TSP tasks. Therefore, we adopted the shortest path obtained from the LKH algorithm as the reference optimal path length for each TSP problem.

## Statistical Analysis of PAO Values and Cross Points

Prior to data analysis, we conducted normality tests on the PAO values and intersection counts across all solutions and individual grouped solutions. The results indicated that none of the datasets adhered to a normal distribution for either PAO values or intersection counts. To evaluate changes in human performance before and after perturbations, we employed the Wilcoxon signed-rank test to compare differences between paired groups. The Wilcoxon signed-rank test(Wilcoxon, 1945, 1947), a non-parametric statistical method, is robust for analyzing differences in paired datasets when normality assumptions are violated, making it suitable for assessing the impact of perturbations on task performance.Furthermore, we computed the effect size of the comparative results using Cohen's d as the metric. Cohen's d measures the magnitude of the difference between the means of two groups relative to the standard deviation of the data. In our analysis, it serves as an effective indicator for quantifying the degree of difference between each pair of groups.

**Line Stylization Degree Differences Across Groups** To verify whether the grouped configurations generated through different perturbations exhibited significant differences in the degree of line stylization, we performed Wilcoxon signed-rank tests on the line stylization scores (Fleiss' Kappa values) across groups. The results indicate that the Fleiss' Kappa values of the problem configurations subjected to random point deletion significantly differ from those of the original configurations ($z = 9.495$, $p < .001$***, $d = 1.136$). Similarly, the configurations subjected to random point displacement also show a significant difference in Fleiss' Kappa values compared to the original configurations ($z = 11.322$, $p < .001$***, $d = 2.028$). Moreover, a significant difference in Fleiss' Kappa values was also observed between the random displacement group and the random deletion group ($z = 11.322$, $p < .001$***, $d = 1.903$). These results indicate that applying distinct perturbations (e.g., point deletion or positional shifting) successfully altered the degree of line stylization in the TSP point arrays. Figure 6a illustrates the distribution of line stylization degree differences between any two groups.

**PAO Values and Cross Points Analysis** After verifying the differences in line stylization degrees across groups, we further conducted Wilcoxon signed-rank tests on the experimental data to analyze PAO values under different line stylization groupings. The results revealed significant differences in PAO values between the original configurations and those perturbed by random point deletion ($z = -1.983$, $p = .047$*,p=-

0.026), as well as between the original configurations and those perturbed by random positional shifting (z = -2.167, p = .03*,d=-0.086) (see Figure 6b for details).

On the other hand, we also analyzed the number of intersections across different groups. The results showed significant differences in the number of intersections between the original configurations and those perturbed by random point deletion (z = -8.557, p <.001***,d=0.814), as well as between the original configurations and those perturbed by random positional shifting (z = -2.071, p = .038*,d=0.232). Additionally, a significant difference was observed between the random positional shifting group and the random point deletion group (z = -7.628, p <.001***,d=-0.701). The specific distribution of these results is illustrated in Figure 6c.

### Interpretation of Results

From the distribution of line stylization degree differences across configurations in Figure 6a, we observe that the Original group exhibits the highest degree of line stylization, the RandomDelete group shows an intermediate level, and the RandomPosition group has the lowest degree. Further Wilcoxon signed-rank test results indicate significant differences in PAO values between the Original group and both the RandomDelete and RandomPosition groups, suggesting that the degree of line stylization significantly influences human efficiency in solving TSP problems. We infer that, in TSP tasks, the line stylization features of point arrays can enhance human path-planning efficiency, particularly when the paths exhibit certain regularities.

On the other hand, the analysis of intersection counts provides another intriguing perspective. While existing studies suggest that a higher number of intersections correlates with lower path-planning efficiency(Rooij et al., 2003; MacGregor et al., 2004), our analysis indicates that the differences in intersection counts are not driven by variations in path quality but rather by the number of points in the problem configurations. In the experimental design, the RandomDelete group perturbed the line stylization degree by deleting 70% of the points, whereas the RandomPosition group perturbed the configurations by shifting point positions, thus maintaining an equal number of points across both groups. The results show that the difference in intersection counts between the Original and RandomPosition groups is less significant, while the differences between these groups and the RandomDelete group are highly significant. This suggests that the number of points is the primary factor influencing intersection counts, rather than the line stylization features themselves.

In summary, our analysis validates the hypothesis that the line stylization features of point arrays in TSP problems significantly influence human problem-solving efficiency. Specifically, when point arrays exhibit stronger line stylization features, participants demonstrate more effective path-planning strategies, reducing unnecessary decision biases and thereby improving TSP-solving efficiency. These findings not only reveal the potential role of visual structures in path planning but also provide novel insights and methodologies for future research on cognitive decision-making.

### Examining the Mediating Effect of Decision-Making Style

Based on the experimental results, we propose a tentative hypothesis: a reduction in the degree of line stylization alters human decision-making strategies, which in turn leads to a decline in performance levels. In this causal chain, decision-making strategies serve as the mediating variable, while the number of intersections acts as a measurable indicator of this mediation. This section aims to investigate whether the independent variable (Fleiss' Kappa) influences the dependent variable (PAO) through the mediating variable (Intersections) and to test the significance of this mediating effect,Fleiss' Kappa corresponds to the degree of line stylization in the problem configurations.

To this end, we employed the PROCESS Macro (Model 4)(Preacher & Hayes, 2004) for mediation analysis(A. F. Hayes, 2009; A. Hayes & Rockwood, 2017) and utilized the Bootstrap method to estimate the confidence intervals of the indirect effects (with 5,000 resampling iterations). Compared to other methods, the Bootstrap method(Baron & Kenny, 1986) does not require assumptions of data normality and provides confidence intervals for the results. The model is shown in Figure 7. The results indicate that the independent variable Fleiss' Kappa has a significant effect on the mediator variable Intersections ($\beta = 5.44$, $p < .001$), and Intersections in turn significantly affects the dependent variable PAO ($\beta = 0.0171$, $p < .001$). However, the direct effect of Fleiss' Kappa on PAO is not significant ($\beta = -0.0526$, $p = 0.105$), suggesting that the influence of Fleiss' Kappa on PAO is primarily mediated through Intersections. The indirect effect, calculated via the bootstrap method, is 0.0933, with a bootstrapped confidence interval of [0.0396, 0.1564], which does not include zero—indicating a significant indirect effect. This confirms Intersections as a mediator between Fleiss' Kappa and PAO, consistent with a full mediation model..

Thus, it can be seen that our hypothesis is well supported, namely, the degree of lineation in TSP configuration fully affects the path length performance in human solutions by influencing human decision-making.

## Discussion and Conclusion

This study investigates the line stylization features of point array distributions in the Traveling Salesman Problem (TSP) and their impact on human path-planning decisions. We proposed a method to quantify line stylization features and experimentally validated their influence on path-planning efficiency across different configurations. The results demonstrate that the line stylization features of point arrays not only affect human decision-making processes but also guide path selection through visual cues, thereby significantly enhancing path-planning efficiency. By comparing TSP configurations with varying degrees of line stylization, we further elucidate how these visual structural characteristics shape cognitive decision-making processes.

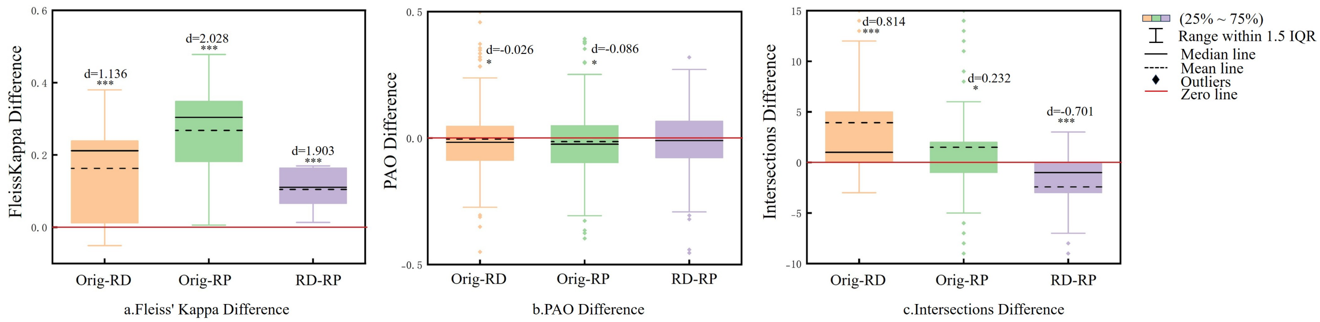


Figure 6: The box plots of the Fleiss Kappa differences, PAO differences, and Intersections differences between the three configurations. The significance symbols above the boxes indicate the level of significance of the differences between the two groups corresponding to each box,*p <.05, ***p <.001,d represents the effect size as measured by Cohen's d.

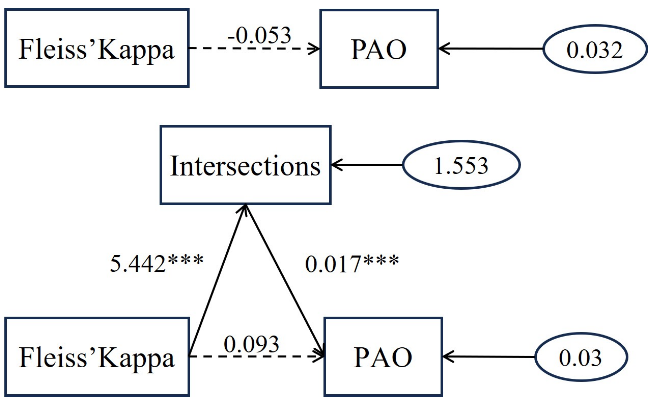


Figure 7: The mediation model diagram. The effect of Fleiss' Kappa on Intersections is significant, the direct effect of Fleiss' Kappa on PAO is not significant, and the indirect effect of Fleiss' Kappa on PAO through Intersections is significant.

The experimental results reveal that when the point arrays in TSP problems exhibit stronger line stylization features, humans demonstrate higher efficiency in path planning. This phenomenon is particularly evident in fewer path intersections and shorter path lengths. Specifically, configurations with higher degrees of line stylization (e.g., the Original group) compared to randomly perturbed configurations (e.g., the RandomDelete and RandomPosition groups) enable humans to more easily identify underlying path structures, leading to more efficient decision-making. In this way, visual structural features provide cognitive cues for path planning, assisting participants in identifying optimal paths within complex problems.

Furthermore, The results indicate that the influence of line stylization features is primarily mediated by adjustments in decision-making strategies. Specifically, line stylization features do not directly impact path-planning quality but instead optimize path selection by influencing participants' cognitive strategies. we conducted a mediation analysis to examine the effects of varying degrees of line stylization on the decision-making process. This finding highlights the critical role of visual structures in path planning, demonstrating that they not only affect decision-making efficiency but also indirectly influence decision quality.

Although this study provides new perspectives on understanding the influence of TSP configurations on path planning, it still has certain limitations. Firstly, the research was conducted solely within a two-dimensional configuration. Future studies could extend this to path planning problems in higher dimensions, exploring how different spatial structures affect the decision-making process. Secondly, while we have quantified the impact of Line Stylization Degree, we have not delved deeply into the underlying cognitive mechanisms.Future research could incorporate techniques such as eye-tracking, the edge detection mechanisms of human vision to conduct experimental studies with a larger pool of participants and analyze the resulting data. Such investigations may further elucidate how the Line Stylization Degree influences the brain's decision-making processes through visual pathways, thereby deepening our understanding of the cognitive mechanisms underlying path planning.

In summary, this study quantifies the Line Stylization Degree of lattice configurations in the TSP problem and reveals how visual structure affects human path planning decisions. The findings of this study hold significant implications for the field of cognitive science. While traditional TSP research has predominantly focused on computational methods and path length optimization, this study highlights how the structural characteristics of TSP configurations—particularly the line stylization features of point arrays—influence human decision-making processes at a visual level. This perspective offers novel insights for cognitive modeling and optimization in path planning and provides valuable reference for understanding how humans utilize visual cues to make decisions in complex tasks. Additionally, the quantification method we proposed introduces a new tool for experimental design in cognitive psychology, enabling future research to more precisely measure and compare the impact of different visual structures on decision-making.

## References


Baron, R. M., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51 6*, 1173-82. doi: 10.1037/0022-3514.51.6.1173

Bickley, S. J., & Torgler, B. (2021). A systematic approach to public health – novel application of the human factors analysis and classification system to public health and covid-19. *Safety Science*, *140*, 105312 - 105312. doi: 10.1016/j.ssci.2021.105312

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37 - 46. doi: 10.1177/001316446002000104

Dry, M. J., Lee, M., Vickers, D., & Hughes, P. (2006). Human performance on visually presented traveling salesperson problems with varying numbers of nodes. *J. Probl. Solving*, *1*. doi: 10.7771/1932-6246.1004

Estévez-Almenzar, M., Baeza-Yates, R., & Castillo, C. (2025). A comparison of human and machine learning errors in face recognition. *arXiv preprint arXiv:2502.11337*.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, *76*(5), 378. doi: 10.1037/h0031619

Fleiss, J. L., Levin, B., Paik, M. C., et al. (1981). The measurement of interrater agreement. *Statistical methods for rates and proportions*, *2*(212-236), 22–23.

Flood, M. M. (1956). The traveling-salesman problem. *Operations Research*, *4*, 61-75. doi: 10.1287/OPRE.4.1.61

Goh, Y. C., Cai, X. Q., Theseira, W., Ko, G., & Khor, K. A. (2020). Evaluating human versus machine learning performance in classifying research abstracts. *Scientometrics*, *125*, 1197–1212.

Graham, S. M., Joshi, A., & Pizlo, Z. (2000). The traveling salesman problem: A hierarchical model. *Memory & Cognition*, *28*, 1191-1204. doi: 10.3758/BF03211820

Hayes, A., & Rockwood, N. J. (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour research and therapy*, *98*, 39-57. doi: 10.1016/j.brat.2016.11.001

Hayes, A. F. (2009). Beyond baron and kenny: Statistical mediation analysis in the new millennium. *Communication monographs*, *76*(4), 408–420. doi: 10.1080/03637750903310360

Helsgaun, K. (2000). An effective implementation of the lin-kernighan traveling salesman heuristic. *Eur. J. Oper. Res.*, *126*, 106-130. doi: 10.1016/S0377-2217(99)00284-2

Kerschke, P., Kotthoff, L., Bossek, J., Hoos, H., & Trautmann, H. (2018). Leveraging tsp solver complementarity through machine learning. *Evolutionary Computation*, *26*, 597-620. doi: 10.1162/evco_a_00215

MacGregor, J. (2015). Effects of cluster location and cluster distribution on performance on the traveling salesman problem. *Attention, Perception, & Psychophysics*, *77*, 2491-2501. doi: 10.3758/s13414-015-0925-2

MacGregor, J., Chronicle, E., & Ormerod, T. (2004). Convex hull or crossing avoidance? solution heuristics in the traveling salesperson problem. *Memory & Cognition*, *32*, 260-270. doi: 10.3758/BF03196857

MacGregor, J., Chronicle, E., & Ormerod, T. (2006). A comparison of heuristic and human performance on open versions of the traveling salesperson problem. *J. Probl. Solving*, *1*. doi: 10.7771/1932-6246.1005

MacGregor, J., & Chu, Y. (2011). Human performance on the traveling salesman and related problems: A review. *J. Probl. Solving*, *3*. doi: 10.7771/1932-6246.1090

MacGregor, J., & Ormerod, T. (1996). Human performance on the traveling salesman problem. *Perception & Psychophysics*, *58*, 527-539. doi: 10.3758/BF03213088

MacGregor, J., Ormerod, T., & Chronicle, E. (1999). Spatial and contextual factors in human performance on the travelling salesperson problem. *Perception*, *28*, 1417 - 1427. doi: 10.1068/p2863

Mueller, P., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard. *Psychological Science*, *25*, 1159 - 1168. doi: 10.1177/0956797614524581

Preacher, K. J., & Hayes, A. (2004). Spss and sas procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717-731. doi: 10.3758/BF03206553

Rach, T., & Kirsch, A. (2016). Modelling human problem solving with data from an online game. *Cognitive Processing*, *17*, 415 - 428. doi: 10.1007/s10339-016-0767-4

Rooij, I., Stege, U., & Schactman, A. (2003). Convex hull and tour crossings in the euclidean traveling salesperson problem: Implications for human performance studies. *Memory & Cognition*, *31*, 215-220. doi: 10.3758/BF03194380

Vickers, D., Lee, M., Dry, M. J., & Hughes, P. (2003). The roles of the convex hull and the number of potential intersections in performance on visually presented traveling salesperson problems. *Memory & Cognition*, *31*, 1094-1104. doi: 10.3758/BF03196130

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 196-202. doi: 10.1007/978-1-4612-4380-9_16

Wilcoxon, F. (1947). Probability tables for individual comparisons by ranking methods. *Biometrics*, *3*, 119. doi: 10.2307/3001946

Zheng, J., He, K., Zhou, J., Jin, Y., & Li, C.-M. (2023). Reinforced lin–kernighan–helsgaun algorithms for the traveling salesman problems. *Knowledge-Based Systems*, *260*, 110144. doi: 10.1016/j.knosys.2022.110144