Title

A Study on the Predictors of Problem Difficulty for the Planning Serious Game Tik Tik

Permalink

https://escholarship.org/uc/item/1vk876q3

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Dimov, Cvetomir

Bavelier, Daphne

Publication Date

2026

A Study on the Predictors of Problem Difficulty for the Planning Serious Game Tik Tik

Cvetomir M. Dimov (cvetomir.dimov@unige.ch) and Daphne Bavelier
Department of Psychology and Educational Sciences, University of Geneva

Abstract

Tik Tik is a serious game designed to study the cognitive mechanisms of planning by combining ecologically valid navigational demands with a highly flexible problem structure. Drawing on data from three studies totaling 122 participants, we show that Tik Tik problem difficulty is primarily driven by a smart, non-exhaustive search of its problem space, in which character trajectories are represented. Furthermore, we find that working memory and visuo-motor skills contribute, highlighting the types of cognitive representations and skills Tik Tik also taps. Our four-parameter model incorporating these factors explains a large proportion of the variance in problem difficulty. Finally, we compare these cognitive demands to those of the most popular planning task, the Tower of London, highlighting commonalities and differences between these two tasks.

Keywords: planning, Tik Tik, structural problem parameters, internal representation, search heuristics

Introduction

Planning is a fundamental cognitive process, essential for navigating the complexities of daily life. This capacity is engaged in a broad spectrum of problems: from mundane routines such as meal preparation (Byrne, 1981) to sophisticated, goal-directed challenges. The latter include classic puzzle-based tasks like the Tower of Hanoi (Simon, 1975) and Water Jug problems (Atwood & Polson, 1976), strategic gameplay in chess (Chase & Simon, 1973) or tic-tac-toe (van Opheusden et al., 2021), and high-stakes professional domains such as preparing for a medical surgery (Xiao et al., 1997).

Given this diversity of settings, it is not surprising that the planning process can recruit distinct cognitive resources and load them differently. Scholnick et al. (1998) illustrated this through a detailed comparison of the processes underlying two popular planning paradigms, the Tower of Hanoi and Errand planning. Indeed, empirical investigations frequently find low correlations between planning tasks (Cunningham et al., 2025), prompting some researchers to question the validity of the construct itself (Morris & Ward, 2004).

We argue, however, that such conclusions are premature. As for every complex cognitive construct (see, for example, executive functions or attention; Miyake et al., 2000), it is likely that, as the domain progresses, different planning constructs can be identified, sharing some common variance. A more granular understanding of the mechanisms involved in planning may clarify why planning outcomes sometimes diverge, specifically by revealing how certain cognitive factors interact to produce these outcomes.

Since Newell and Simon (1972)'s human problem solving theory, it is well accepted that problem solving (and the planning it entails) involves two subprocesses: forming an internal representation of the problem, the *problem space*, and *searching* in that problem space. The problem space includes start and goal states, intermediate states, transitions between these states, and task constraints (Bassok & Novick, 2012). The problem space is typically visualized with a graph, in which states are represented as nodes and possible transitions between these states as edges. Search in the problem space can take place both physically, by executing actions, or mentally, through processes such as planning. Planning mentally traverses the problem space until a suitable plan that reaches the goal state is found.

Problem solving tasks can differ in both their representational and search demands, which in turn influence task difficulty. To inform us about the demands of a planning task, its structural problem parameters need to be systematically examined. For example, the most widely used planning research tool, the Tower of London (ToL; Shallice, 1982) was described in terms of eight parameters (Berg et al., 2010; Ward and Allport, 1997; see Kaller et al., 2011, for an overview). As predicted by the problem space theory, some of these parameters (the *minimum number of moves*, the *number of optimal paths to solution* and the *number and type of suboptimal solution paths*) were measures of the size of the problem space. Additional parameters revealed how the search process unfolded. For example, problems with *global-local conflict moves* were more difficult than problems without, indicating a general bias against locally moving away from the goal state.

The present work introduces Tik Tik (Dimov & Bavelier, 2025), a novel planning task designed to tap cognitive processes similar to those of the ToL. Tik Tik utilizes an ecologically valid navigation paradigm (Mattar & Lengyel, 2022) and offers significant flexibility in problem structure, allowing for precise and wide-ranging variations in complexity. Furthermore, its gamified design provides high motivational value for participants. By verifying the key parameters that drive difficulty in Tik Tik, we aim to situate its cognitive demands within the broader landscape of other planning tasks.

Before we engage in the modeling investigation, we will review Tik Tik and outline its primary strengths as a research tool. To identify the cognitive structure of Tik Tik, we will start by generating a comprehensive list of candidate predictors based on hypotheses about what representation is formed

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

and how planning unfolds specifically in this task. We will then describe the data set that we fit the model to, which is compiled from three studies. Finally, we present a 4-parameter model that significantly predicts subject success. We will conclude by discussing the implications of our model about the underlying internal representation and search process, and compare these to the Tower of London.

Tik Tik

Tik Tik is a serious game specifically designed to facilitate the empirical study of planning. The objective is to navigate two characters, a fire ball and an ice cube, to their respective target locations, indicated by an empty fire circle and an empty ice square (see Figure 1 for an example problem). To reach these targets, players must strategically navigate two types of obstacles: fire rays and ice rays.

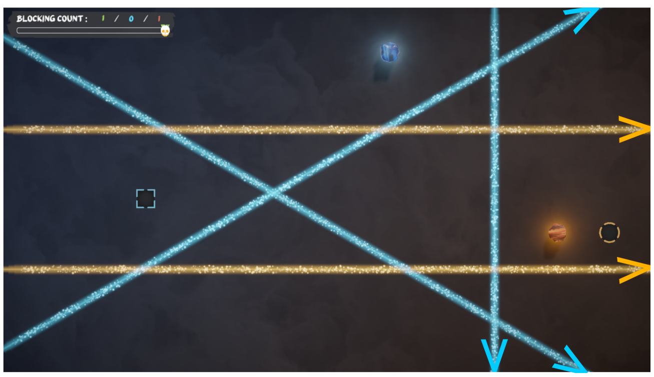

Figure 1: Tik Tik exemplified via one problem in the serious game. A fire ball and an ice cube have to be navigated to target fire and ice locations. Fire and ice rays need to be blocked to reach these locations. The ray flow direction is indicated with superimposed arrows. The top left shows the blockings count (in blue), the minimum number of blockings possible (in green) and the blockings limt (in red).

The characters possess complementary abilities and vulnerabilities. The fire ball is destroyed upon contact with an ice ray but can freely traverse or block fire rays. Conversely, the ice cube must avoid fire rays but can traverse or block ice rays. Blocking occurs when a character solidifies while standing in the path of a ray of its own element. By doing so, it cuts off the ray in the direction of flow and creates a safe passage for the other character. A problem is successfully solved when the player discovers a sequence of blockings that allows both characters to reach their targets.

In Tik Tik, blockings serve as the primary "currency" of gameplay. The interface displays the current number of blockings used (Figure 1, in blue), the minimum number required for an optimal solution (in green), and a maximum blocking limit (in red). By adjusting this limit and providing incentives for achieving the minimum-blockings solution, we can incentivize the players to pursue optimality, increasing the effort they invest in planning.

While Tik Tik was designed to ensure an engaging user experience (see Dimov and Bavelier, 2025, for design details), its primary advantage for research lies in its scalability within an ecologically valid framework - a need recently highlighted in the literature (Mattar & Lengyel, 2022). We can generate a vast array of problems with precisely controlled levels of complexity. Specifically, we can directly manipulate the problem space by varying the number and configuration of rays, relying on the fact that each additional ray partitions the playing field into distinct regions, or states, increasing a problem's complexity.

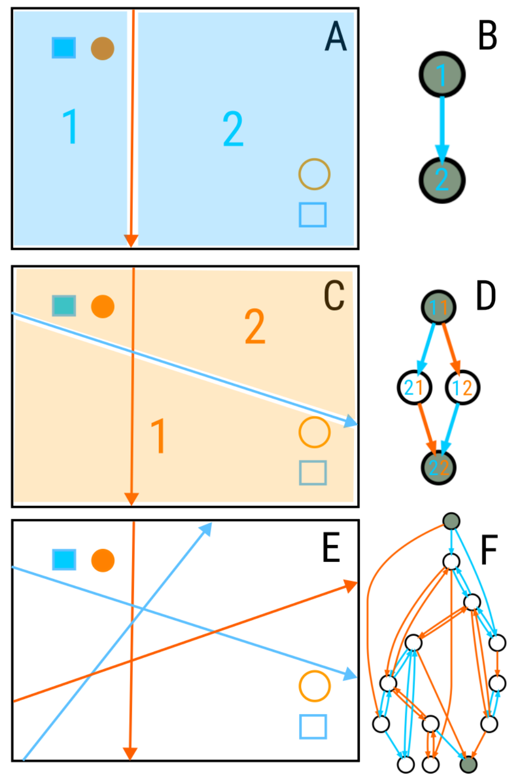

Figure 2: A-B. Problem sketch and problem space graph of a problem with just a single fire ray. The fire ray splits the field into two ice cube regions, or problem states. C-D. Problem sketch and problem space graph of a problem with one ice and one fire ray. Each ray creates two regions for each of the characters, for a total of 2x2 = 4 possible states. E-F. Problem sketch and problem space graph of a problem with two ice and two fire rays in a particular configuration. 13 different states can be reached from the start state.

Let us illustrate how with an example. Perhaps the simplest level is when a single ray (e.g., fire ray) bisects the field into two regions (Figure 2, A). Since the ice cube cannot cross this ray freely, these regions represent two discrete states between which it can only transition when the fire ray is blocked. Consequently, the problem space graph of such a problem is represented with two nodes, representing the two ice cube states, which are linked by a blue arrow, representing an ice cube transition (Figure 2, B). Note that a transition between two states is a complex action, consisting of blocking a ray with one character (here, the fire ball) and navigating to the new state with the other (here, the ice cube).

Now, adding an ice ray creates a new problem space, partitioning the field also for the fire ball (Figure 2 C). Since a state is the combination of the individual characters' states, the total state count is the product of the individual state counts. Specifically in this configuration there are 2 ice states × 2 fire states = 4 total states (see Figure 2, D). We can continue adding rays to increase problem complexity (e.g., Figure 2, E & F). Because each new ray increases the size of the problem space and because we can arrange their configuration and character start and goal locations as we desire, Tik Tik allows for the creation of problems ranging from the trivial to the very hard.

Candidate predictors

We identified seven candidate structural parameters to explain variation in problem difficulty, drawing on established findings from the planning and working memory literatures (see Table 1 for an overview). In line with Newell and Simon (1972)'s formalism, these variables are categorized into three theoretical domains: Search (reflecting search strategy and graph structure), Representation and Working Memory (reflecting cognitive capacity and internal representation of the task), and Motor Execution (reflecting motor precision demands).

Search Strategy

Humans rarely explore a problem space exhaustively (Morris & Ward, 2004). Instead, they utilize heuristics to prune the problem space and focus on a promising subset of states. This principle of strategic, frugal search is well-documented across planning tasks ranging from errand planning (Gärling, 1996) to probabilistic tasks (Huys et al., 2012) and complex board games (van Opheusden et al., 2021). The four variables *Depth*, *Breadth*, *Single-initial*, and *Blocking-repeat* operationalize hypotheses about how problem complexity and search strategy interact.

Search Depth and Breadth. The most fundamental predictor of problem difficulty is the size of the problem space that is being searched. We hypothesize that in Tik Tik search is limited to a depth equal to the minimum number of blockings. Since the objective is to find the solution with the minimum number, we expect that subjects would abort search if it does not reach the goal state in that number and move to another candidate solution. This effectively truncates the problem space graph, producing a graph of *Depth* equal to the minimum number of blockings and of *Breadth* equal to the number of paths up to that depth (see Figure 3, A).

Single Initial Transition (*Single-initial*). The variable *Single-initial* characterizes the structure of the root of the search graph. If the problem space permits only one viable initial transition, *Single-initial* = 1, whereas if multiple transitions are possible, *Single-initial* = 0. We hypothesize that problems where *Single-initial* = 1 allow planners to "offload" the first step via immediate execution. This effectively reduces the remaining search depth by one transition, thereby decreasing cognitive demand.

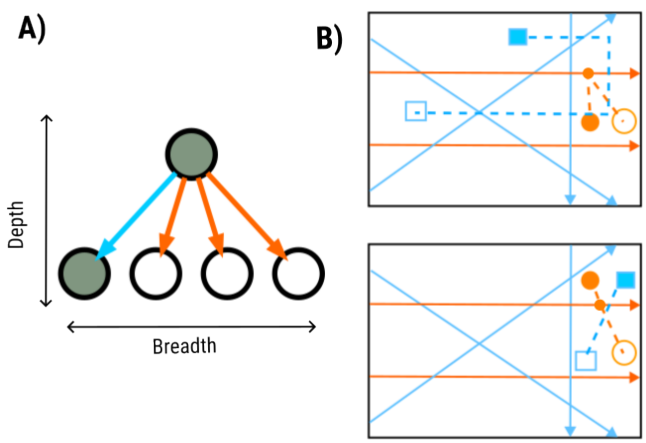

Figure 3: A. Problem space graph. The Depth of the graph equals to the minimum number of blockings, whereas the Breadth the total number of paths of length Depth. B. Number of straight trajectory segments. Both examples have the same ray configuration and require a single blocking. In the top example, the ice cube must follow a complex trajectory consisting of three straights segments and the fire ball a trajectory that consists of two (5 segments in total), because of their specific start and goal locations. The bottom examples allows each of two characters to follow a straight path, for a total of 2 straight segments. The search graph on panel A corresponds to both of these problems.

Repetitive Blocking (*Blocking-repeat*). *Blocking-repeat* indicates whether the optimal solution requires a single ray segment to be blocked more than once (*Blocking-repeat* = 1). This variable captures a known cognitive resistance to backtracking or moving away from a perceived goal state. For example, in the Tower of Hanoi, subjects rarely choose to reverse a previous move (Anzai & Simon, 1979). Similarly, Tower of London problems involving global-local conflicts are significantly more difficult than matching problems that involve no such conflicts (Kaller et al., 2011). In Tik Tik, we expect that subjects will be less likely to consider blocking again a ray that they already blocked in their plan.

Representation and Working Memory

Any problem needs to be internally represented before planning commences (Newell & Simon, 1972). Problem representation can have a profound impact on problem difficulty, as indicated by research with isomorphs (Kotovsky et al., 1985) and on insight problem solving (Öllinger & Knoblich, 2009). Two candidate variables reveal how the internal representation

Table 1: Summary of Candidate Structural Predictors for Problem Difficulty in Tik Tik

Category	Parameter	Description	Theoretical Basis
Search	*Depth*	Minimum number of blockings in problem	Search depth (Newell & Simon, 1972)
	Breadth	Number of paths of optimal length	Search breadth (Newell & Simon, 1972)
	Single-initial	Presence of a single initial move (binary)	Heuristic strategy (Morris & Ward, 2004)
	Blocking-repeat	Ray segment blocked multiple times (binary)	Avoidance of backtracking (Anzai & Simon, 1979)
Representation & WM	N_{straight}	Number of distinct linear path segments	Visuospatial WM load (Kemps, 2001)
	N_{switch}	Frequency of blocking character switches	Object switch costs (Oberauer et al., 2013)
Motor	N_{narrow}	Count of spatial constraints/tight gaps	Sensorimotor precision demands

of the problem interacts with limitations on working memory (WM).

Character Switch Costs (N_{switch}). N_{switch} quantifies the number of times a player must switch between blocking with the fire ball and blocking with the ice cube within the optimal solution. Optimal solutions of the same length can incurr very different character switch counts. For instance, the 3-blockings sequence in the order Fire → Ice → Fire involves two switches, whereas a sequence of only Fire blockings involves zero. Drawing on the object switch cost literature (Garavan, 1998; Oberauer, 2009), we theorize that maintaining and updating the active character in WM imposes a cognitive load that is independent of the total number of blockings performed.

Number of Straight Trajectory Segments (N_{straight}). N_{straight} quantifies the geometric complexity of the solution path. It is calculated by tracing the trajectories of both characters in the optimal solution and counting the number of distinct linear segments. We operationalized this using a 30-degree threshold: a trajectory is considered as consisting of a single straight segment until a directional change exceeds 30 degrees. This applies as one of the characters moves after the other has blocked a ray and freed the way. Such a path may require traversing several straight segments if the path is winding (for example, see the ice cube's trajectory in Figure 3, B, upper panel). This also applies to the trajectories that characters follow from one blocking to another and to the final goal position (for example, see fire ball path in Figure 3, B, upper panel).

Theoretically, N_{straight} serves as a proxy for visuospatial WM load, as path complexity is known to influence performance in visuospatial WM (Kemps, 2001) and navigation tasks (Yesiltepe et al., 2023). If significant, this parameter would indicate that subjects do not remember only which ray to block (as indicated by *Depth*), but also the entire path to reach that ray and the goal location. Consequently, levels with the same graph and even the same ray configuration, but different path complexities may not be equally easy (e.g., compare the two panels in Figure 3, B).

Motor Execution

Number of Narrow Regions (N_{narrow}). The final parameter, the number of narrow regions N_{narrow}, quantifies the motor precision required by the solution. We define a narrow region as any spatial constraint smaller than twice the character's width, such as gaps between lethal rays (i.e., rays of the opposite material than the character) or high-precision blocking zones (i.e., zones which risk blocking the wrong ray). These areas require precise sensorimotor control to avoid mistakes. We hypothesize that a higher N_{narrow} increases the probability of motor execution errors, independent of the player's planning accuracy. This parameter is both a nuisance parameter that should be kept to its minimum when it comes to capturing problem solving constructs as well as a sanity check for the visuo-motor demands of our task.

A Modeling Investigation of Tik Tik's Problem Structure

The Model Fitting Dataset

We developed our model using a pooled dataset derived from three separate studies run on Prolific. Across all studies, participants followed a standardized protocol: they first completed a task tutorial before proceeding to the main problem sets. Problems were designed to be of reasonable difficulty and to have just one optimal solution, consistent with other planning problem sets (Kaller et al., 2012). Problems were presented in a fixed order of hypothesized increasing difficulty. The exact same order was used for all subjects in a study. To maintain high engagement, participants were incentivized with performance-contingent bonuses and provided with immediate feedback upon problem completion. Consistent with other planning studies, participants worked under a designated time limit (Kaller et al., 2012).

Each study was designed with specific procedural and theoretical objectives. To summarize, Study 1 demonstrated that the strictest blocking limit (i.e., equal to the minimum number of blockings) would increase planning compared to a more relaxed limit. Studies 2 and 3 adopted the strict blocking limit from Study 1. Study 2 failed to find evidence about the necessity of warm-up problems, indicating that the tutorial was sufficient for initial practice. It also expanded the range of problem complexity on the upper end of the spectrum. Finally, Study 3 utilized a within-subject design to test two structural parameters and also introduced a procedural change, reducing the time limit from 180 to 120 seconds. The detailed results of these studies are beyond the scope of the current work.

The combined dataset included 122 subjects (81 male; M_{age} = 38 years, range: 18–80) and 66 unique problems, representing a wide variance across the seven candidate parameters. Specifically, the problems encompassed values for *Depth* in the range $[1, 4]$, *Breadth* within $[1, 26]$, N_{straight} within $[2, 8]$, N_{switch} within $[0, 3]$, and N_{narrow} within $[0, 4]$. Both binary variables (*Single-initial* and *Blocking-repeat*) had multiple problems for each value.

Our primary dependent variable was the success rate (solved vs. unsolved). Because Study 3 utilized a shorter time limit (2 minutes) than Studies 1 and 2 (3 minutes), we standardized the dataset to equate the cognitive challenge. Any problems solved after 120 seconds in Studies 1 and 2 were recoded as "unsolved." This was a conservative adjustment, affecting only 1.9% of solutions in Study 1 and 1.3% in Study 2.

Analysis Software

All data analysis was performed with the `R programming language` (R Core Team, 2026) with the `tidyverse` set of packages (Wickham et al., 2019), including `ggplot` for plotting. The logistic regression was fitted with the `lme4` package (Bates et al., 2015), marginal and conditional R^2 were computed with package `MuMIn` (Bartoń, 2026), and VIF (variance inflation factor) was computed with package `performance` (Lüdecke et al., 2021).

Modeling Results

We employed a mixed-effects logistic regression to quantify difficulty as a function of the seven parameters. This approach accounted for the binary outcome variable while treating subject-level differences in planning ability as random effects. We fitted the model to each study individually and to the combined dataset. Model fit was evaluated using Nakagawa and Schielzeth's pseudo-R2 for generalized linear mixed models (theoretical method). In this initial analysis, *Depth*, *Breadth*, *Single-initial*, and N_{straight} emerged as significant cognitive predictors across the three studies and highly significant in the pooled dataset (all $p < 0.001$). The cognitive *Blocking-repeat* was significant only in Study 3 ($p < 0.05$), whereas N_{switch} did not reach significance in any of the studies. Finally, the motor variable N_{narrow} was significant in Studies 1 and 2, and highly significant in the pooled dataset ($p < 0.001$).

While the initial model effectively predicted outcomes, we identified N_{narrow} as a conceptual confound. Because N_{narrow} measures motor precision rather than cognitive planning, difficulty arising from this parameter do not reflect the planning algorithm we aim to isolate. Furthermore, N_{narrow} is a relatively coarse metric, because it treats all narrow regions as

Table 2: Results of multiple logistic regression model for three studies separately and together.

	Study 1	Study 2	Study 3	All
Depth	-0.58^{***}	-0.38^{**}	n.s.	-0.47^{***}
Breadth	-0.08^{+}	-0.10^{**}	-0.11^{*}	-0.09^{***}
Single-initial	1.14^{*}	1.31^{**}	n.s.	0.65^{***}
Blocking-repeat	n.s.	n.s.	n.s.	n.s.
N_{switch}	n.s.	n.s.	n.s.	n.s.
$N_{straight}$	-0.42^{***}	-0.41^{***}	-0.51^{***}	-0.44^{***}
Marginal R^2	0.28	0.42	0.16	0.29

equivalent. In doing so, it ignores factors like the orientation of the gap relative to the cardinal-direction controls, how narrow the regions are, or subject motor ability. To isolate the cognitive components of planning, we excluded all problems containing narrow regions ($N_{\text{narrow}} > 0$) and re-ran the analysis. In this refined dataset, the same four non-motor variables emerged as the primary drivers of difficulty: *Depth*, *Breadth*, *Single-initial*, and N_{straight}, all highly significant ($p < .001$; see Table 2). The fixed effects accounted for a substantial proportion of the variance in the outcome (*Marginal* $R^2 = 0.29$), whereas the random effects did not explain additional variance (*Conditional* $R^2 = 0.29$). The final model did an excellent job of capturing problem solution rate (Figure 4).

Three of four significant variables correlated moderately ($r_{\textit{Depth-Breadth}} = 0.55$, $r_{Depth-N_{straight}} = 0.48$, $r_{\textit{Breadth}-N_{straight}} = 0.55$). Yet, the variance inflation factor ($\text{VIF}_{Depth} = 1.47$, $\text{VIF}_{Breadth} = 1.65$, $\text{VIF}_{N_{straight}} = 1.23$, and $\text{VIF}_{Single-initial} = 1.30$) indicated that there are no multicolinearity issues with the model fit.

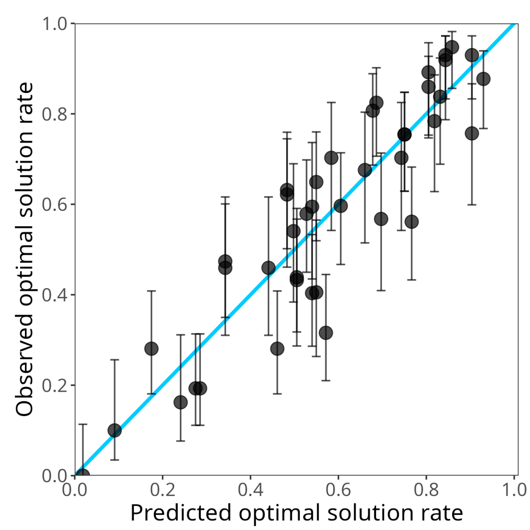

Figure 4: Calibration plot of structural problem parameter model. X-axis: model fit of problem solution rate. Y-axis: empirical solution rate.

Discussion

Tik Tik was developed as a serious game to study human planning across the problem complexity spectrum in an engaging environment. Our findings identify four core cognitive parameters and one motor-based parameter that together account for the vast majority of variance in performance. In the following sections, we discuss the planning process suggested by our model.

Internal Representation

Since Newell and Simon (1972), problem solving has been framed as a search within a problem space - an internal representation of the task. Consequently, the internal representation is the foundation for the more specific planning process. Our results for *Depth* indicate that each blocking action is explicitly represented. However, the significance of N_{straight} indicates that the representation extends beyond a simple list of discrete actions. This measure of path complexity suggests that the internal model of a solution involves the specific trajectories characters must follow. Simpler, more linear paths likely require less information to be maintained in visuospatial working memory. This interpretation is consistent with established path-complexity effects in other tasks, such as the Corsi block-tapping task, where configural aspects of a path significantly impact performance (Kemps, 2001; Parmentier et al., 2005).

Interestingly, there was no indication that higher N_{switch} makes a problem more difficult, contrary to findings in the WM literature regarding object switch costs. This may be a consequence of the task's execution demands: each blocking by one character is followed by navigating the other character to a new region. Consequently, regardless of whether the same character performs consecutive blockings, there is a constant demand to switch focus between the ice cube and fire ball during the execution phase.

Search Strategy

Depth and *Breadth* indicate that the complexity of specific sub-regions of the problem space dictates difficulty. Specifically, these parameters suggest that subjects do not engage in exhaustive search. Rather, they explore only the portion of the space necessary to secure a high probability of success (as argued by Morris and Ward, 2004). In Tik Tik, where the goal is to find the path requiring the fewest blockings, subjects appear to truncate the problem space graph at the depth of the optimal solution. This pattern of non-exhaustive search aligns with findings in various planning tasks, from puzzle-based tasks to complex games (Greeno, 1974; van Opheusden et al., 2017).

The *Single-initial* parameter further points toward a heuristic search strategy. When the cognitive system identifies only one viable immediate action, that action can be executed immediately, allowing the formal search process to commence from the second state. This effectively reduces the problem's cognitive load to that of a simpler task with one transition less. Such a heuristic is aligned with other strategies for simplifying planning, such as pruning (Huys et al., 2012, 2015).

Finally, contrary to findings with other planning tasks, we found no significant effect of our measure of backtracking, *Blocking-repeat*. This might indicate that there is no bias against returning to a previous state in our task. Alternatively, we believe that the effect of this variable might be captured by N_{straight}, because trajectories that require returning back to a previously blocked ray typically consist of more segments.

Motor Execution

Methodologically, identifying N_{narrow} as a significant predictor was a pivotal step. It served as a sanity check aligning with the visuomotor demands of the task. Yet, from the perspective of planning research, variance explained by motor skill is effectively measurement noise. Our principled decision to exclude problems containing narrow regions was therefore essential for isolating the cognitive planning algorithm from sensorimotor execution errors. That we identified the same cognitive parameters after excluding N_{narrow} in facts lends more credence to their centrality in predicting planning difficulty.

Comparing the Cognitive Demands of Tik Tik and the Tower of London

Collectively, our four-parameter model suggests that problem difficulty in Tik Tik is multifaceted, arising from the interplay of problem complexity (*Depth* and *Breadth*), search heuristics (*Single-initial*), and the visuospatial encoding of trajectories (N_{straight}). How do these demands align with those of the Tower of London (ToL), the field's most established planning paradigm?

Kaller et al. (2011) identified eight parameters predicting ToL difficulty, three of which directly correspond to our complexity metrics: the minimum number of moves (equivalent to *Depth*), the number of suboptimal solution paths (equal to *Breadth*-1), and the number of optimal solution paths. This convergence reinforces Newell and Simon (1972)'s formalism, confirming that the problem space structure remains a universal predictor of planning complexity across disparate environments. Unlike broader ToL problem sets, our study utilized problems with unique solutions, mirroring the canonical Freiburg problem set (ToL-F; Kaller et al., 2012). Yet, we expect the presence of multiple optimal solution to reduce difficulty in Tik Tik as well by increasing the probability of success.

The most striking difference is the role of trajectory complexity (N_{straight}). While ToL focuses on the discrete relocation of objects, Tik Tik requires the mental simulation of continuous paths, imposing a unique visuospatial load. This comparison affirms that planning is not a monolithic construct, but a flexible recruitment of cognitive resources that shift according to the environment's structural constraints.

Acknowledgments

The research is funded by Swiss National Science Foundation grant PZ00P1_201837 *Constructing an integrated model of language use during collaborative skill acquisition* awarded to Cvetomir Dimov. We also thank the Klaus J. Jacobs Foundation for their funding of Daphne Bavelier.

References

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review*, *86*(2), 124–140. https://doi.org/10.1037/0033-295x.86.2.124

Atwood, M. E., & Polson, P. G. (1976). A process model for water jug problems. *Cognitive Psychology*, *8*(2), 191–216. https://doi.org/10.1016/0010-0285(76)90023-2

Bartoń, K. (2026). *Mumin: Multi-model inference* [R package version 1.48.19]. https://CRAN.R-project.org/package=MuMIn

Bassok, M., & Novick, L. R. (2012). Problem solving. In *The oxford handbook of thinking and reasoning* (pp. 413–432). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0021

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berg, W. K., Byrd, D. L., McNamara, J. P., & Case, K. (2010). Deconstructing the tower: Parameters and predictors of problem difficulty on the tower of london task. *Brain and Cognition*, *72*(3), 472–482. https://doi.org/10.1016/j.bandc.2010.01.002

Byrne, R. (1981). Mental cookery: An illustration of fact retrieval from plans. *The Quarterly Journal of Experimental Psychology Section A*, *33*(1), 31–37. https://doi.org/10.1080/14640748108400766

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Cunningham, E. G., Bavelier, D., & Green, C. S. (2025). Rethinking planning metrics: An analysis of common measurements of planning abilities. *Cognition*, *263*, 106220. https://doi.org/10.1016/j.cognition.2025.106220

Dimov, C. M., & Bavelier, D. (2025). Tik-tik: A video game for the study of search in problem solving. In *Serious games* (pp. 189–205). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-10518-9_17

Garavan, H. (1998). Serial attention within working memory. *Memory Cognition*, *26*(2), 263–276. https://doi.org/10.3758/bf03201138

Gärling, T. (1996). Sequencing actions: An information-search study of tradeoffs of priorities against spatiotemporal constraints. *Scandinavian Journal of Psychology*, *37*(3), 282–293. https://doi.org/10.1111/j.1467-9450.1996.tb00660.x

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology*, *6*(2), 270–292. https://doi.org/10.1016/0010-0285(74)90014-0

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: How the pavlovian system sculpts goal-directed choices by pruning decision trees (L. T. Maloney, Ed.). *PLoS Computational Biology*, *8*(3), e1002410. https://doi.org/10.1371/journal.pcbi.1002410

Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., Dayan, P., & Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, *112*(10), 3098–3103. https://doi.org/10.1073/pnas.1414219112

Kaller, C. P., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural Brain Research*, *216*(1), 1–8. https://doi.org/10.1016/j.bbr.2010.07.029

Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the tower of london task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment*, *24*(1), 46–53. https://doi.org/10.1037/a0025174

Kemps, E. (2001). Complexity effects in visuo-spatial working memory: Implications for the role of long-term memory. *Memory*, *9*(1), 13–27. https://doi.org/10.1080/09658210042000012

Kotovsky, K., Hayes, J., & Simon, H. (1985). Why are some problems hard? evidence from tower of hanoi. *Cognitive Psychology*, *17*(2), 248–294. https://doi.org/10.1016/0010-0285(85)90009-x

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Mattar, M. G., & Lengyel, M. (2022). Planning in the brain. *Neuron*, *110*(6), 914–934. https://doi.org/10.1016/j.neuron.2021.12.018

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Morris, R., & Ward, G. (2004). *The cognitive psychology of planning*. Psychology Press. https://doi.org/10.4324/9780203493564

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Echo Point Books; Media.

Oberauer, K. (2009). Chapter 2 design for a working memory. In *The psychology of learning and motivation* (pp. 45–100). Elsevier. https://doi.org/10.1016/s0079-7421(09)51002-x

Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments and a computational model. *Cognitive Psychology*, *66*(2), 157–211. https://doi.org/10.1016/j.cogpsych.2012.11.001

Öllinger, M., & Knoblich, G. (2009). Psychological research on insight problem solving. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85198-1_14

Parmentier, F. B. R., Elford, G., & Maybery, M. (2005). Transitional information in spatial serial memory: Path characteristics affect recall performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 412–427. https://doi.org/10.1037/0278-7393.31.3.412

R Core Team. (2026). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org

Scholnick, E. K., Friedman, S. L., & Wallner-Allen, K. E. (1998). What do they really measure? a comparative analysis of planning tasks. In *The developmental psychology of planning*. Psychology Press.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *298*(1089), 199–209. https://doi.org/10.1098/rstb.1982.0082

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*(2), 268–288. https://doi.org/10.1016/0010-0285(75)90012-2

van Opheusden, B., Galbiati, G., Kuperwajs, I., Bnaya, Z., li, Y., & Ma, W. J. (2021). *Revealing the impact of expertise on human planning with a two-player board game*. https://doi.org/10.31234/osf.io/rhq5j

van Opheusden, B., Gilbiati, G., Bnaya, Z., Li, Y., & Ma, W. J. (2017). A computational model for decision tree search. *39th Annual Meeting of the Cognitive Science Society*, 1254–1259. http://refhub.elsevier.com/S2352-1546(19)30062-2/sbref0175

Ward, G., & Allport, A. (1997). Planning and problem solving using the five disc tower of london task. *The Quarterly Journal of Experimental Psychology Section A*, *50*(1), 49–78. https://doi.org/10.1080/713755681

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Xiao, Y., Milgram, P., & Doyle, D. (1997). Planning behavior and its functional role in interactions with complex systems. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *27*(3), 313–324. https://doi.org/10.1109/3468.568740

Yesiltepe, D., Fernández Velasco, P., Coutrot, A., Ozbil Torun, A., Wiener, J., Holscher, C., Hornberger, M., Conroy Dalton, R., & Spiers, H. (2023). Entropy and a sub-group of geometric measures of paths predict the navigability of an environment. *Cognition*, *236*, 105443. https://doi.org/10.1016/j.cognition.2023.105443