Gestalt Effects in Planning:
Rush-Hour as an example

Stefano Bennati (bennati@cognition.uni-freiburg.de)
Sven Brüssow (sven@cognition.uni-freiburg.de)
Marco Ragni (ragni@cognition.uni-freiburg.de)
Lars Konieczny (lars@cognition.uni-freiburg.de)
Center for Cognitive Science, University of Freiburg

Abstract

Planning problems have been extensively studied with regard to graph theoretical properties such as the number of steps necessary to reach a specific goal state or the size of the problem space. These structural properties, however, do not completely characterize a problem. In the presented eye-tracking study we also investigated the influence of perceptual factors on the solution to a planning problem. While not affecting the correctness of a solution, the results suggest that certain Gestalt properties are responsible for the deviation from optimal plans.

Keywords: Move planning; Rush-Hour; Gestalt

Introduction

Planning problems can be characterized by structural properties, such as the number of steps necessary to reach a specific goal state or the size of the problem space, and perceptual properties, such as colors and spatial relations between elements. While structural elements have been widely studied, the latter have not receive as much attention. The reason may be that structural properties are easier to manipulate than perceptual properties. The problem of our choice, Rush-Hour, can be easily manipulated with regard to perceptual properties.

Rush-Hour schematizes a crowded parking lot on a 6×6 grid (cf. Fig. 1) and the task is to clear the way for the player's car which is blocked by some other vehicles. The player's car is always red, horizontally aligned and placed in the third row, the same row where the exit is. There are two types of vehicles: cars (length two) and trucks (length three). Each vehicle has an orientation–vertical or horizontal–and a color. All vehicles can only be moved forward and backward along their longitudinal axes. The game rules forbid moving a vehicle over or through another vehicle or breaking the walls that surround the parking lot. The goal is to clear the way to the exit by sequentially moving the vehicles that block the way, which are in turn blocked by others.

The game is well-defined, decomposable, non dynamic and has only one goal. It is also PSPACE-complete (Flake & Baum, 2002). There is normally more than one possible solution, but only few of them are optimal. We define an *optimal solution* as the solution that involves the least possible number of moves.

Planning problems are often characterized by permutation problems. From a cognitive perspective planning can be defined as the anticipation of action steps or "a procedure for achieving a particular goal or desired outcome" (Morris & Ward, 2005).

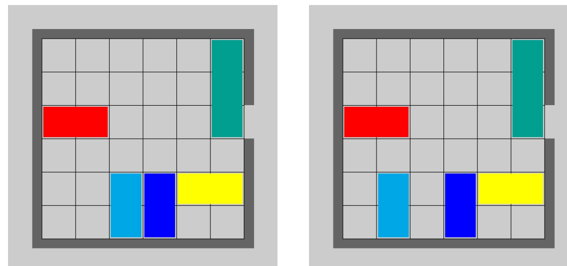


Figure 1: Rush-Hour sample configurations. The task is to rearrange the vehicles such that the red car can be moved out. Only the board on the left contains a cluster.

Insights from different domains, such as Tower of London, indicate that difficulties arise from static properties such as planning depth, i.e., the number of moves necessary to transform the initial state into a goal state (Kaller, Unterrainer, Rahm, & Halsband, 2004; Kaller, Rahm, Köstering, & Unterrainer, 2011).

A relevant property of Rush-Hour is the number of vehicles on the grid, as it increases the search tree. Dynamic properties, such as the number of counter-intuitive moves and the number of circular move sequences have an influence on difficulty as well (Ragni, Steffenhagen, & Fangmeier, 2011). A move is counter-intuitive if it results in a higher distance to the goal state, for example when the goal car has to be moved away from the exit or into an exit-blocking position. A move sequence is circular if one of its vehicles is blocked by another vehicle that is part of the same move sequence.

Human performance on two different boards with the same structural properties may differ. We propose that this difference results from perceptual properties that affect problem solving and planning processes in a way that deserves deeper investigation.

Gestalt Theory asserts that under certain conditions people perceive a group of distinct objects as a holistic unit. For example, when objects are aligned to each other or are in close proximity to each other, they will be perceived as parts of a bigger object (Koffka, 1935; Köhler, 1959; Wertheimer, 1938). With respect to the Rush-Hour domain, a *cluster* represents a meta-object that groups together adjacent objects on the board: It is defined as a group of two or more vehicles that are next to each other such that their major axes are parallel (cf. Fig. 2). In this context, the Gestalt laws of proximity and

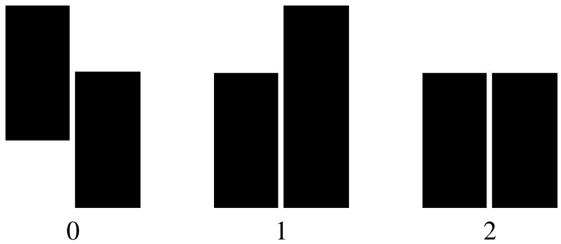


Figure 2: Gestalt levels. A cluster of two cars can be categorized based on how many of their endings are aligned.

continuity can be applied. Continuity states that elements of objects tend to be grouped together if they are aligned within a object.

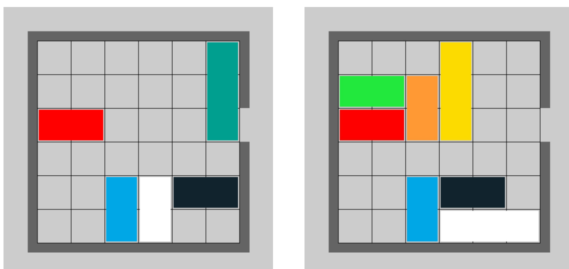


Figure 3: Examples of clusters containing two adjacent cars forming a Level 2 (left, blue and white) and a Level 1 distracting cluster (right, black and white).

An example configuration is shown in Fig. 3 (left) that contains two adjacent cars forming a cluster (Level 2). Both elements cross the axis of the black car, which is indirectly occluding the exit. The black car needs to be moved one step left to unblock the previous element of the move sequence. The closest element of the cluster, the white car, is just beside it, the other element one cell farther away. To solve the problem, only the white car needs to be shifted to make room for the black one, the other car can stay at its place. The peculiarity of such a configuration is that choosing to shift both elements of the cluster does not preclude the goal state, it only requires one more step. The example in Fig. 3 (right) shows a distracting cluster (Level 1). In this case the cluster (bottom right) must be separated to solve the problem optimally: The black car in the cluster has to be moved to the right and the white truck on the bottom has to be moved to the left. Another possible solution would allow to move it left as well and keep the cluster together, at the price of an extra solution step.

A distracting cluster must be split-up in order to optimally solve the board: At least one vehicle (the distracting element) must be moved in a different direction or by a different distance than the others (possibly zero). A suboptimal solution might involve uniform movement of all elements in the cluster; as a result the configuration of the cluster will remain unchanged. The distracting cluster forms a meta-object that attracts attention and requires mental effort when its components have to be separated. It seems reasonable to assume that players split up the distracting cluster only if necessary, therefore they more often arrive at a suboptimal solution that moves the cluster as a unique entity and requires a greater number of moves. Our hypothesis is that participants do not solve the game by a simple depth search rather, they are influenced by the visual representation of the game, especially by distracting clusters, meta-objects that attract attention and require mental effort when they are separated into their components.

Methods

Participants

Thirty participants (15 female, $M = 28.27$ years, $SD = 7.69$) recruited at the University of Freiburg, Germany, took part in this study. Participants gave informed consent and were either paid or received course credit for their participation.

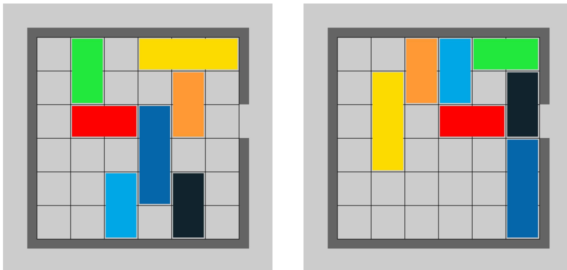
Procedure, materials, and design

Participants were seated in front of a 20-inch flat screen with a resolution of 1600×1200 pixels. Eye movements were recorded using an EyeLink 1000 remote system (SR Research), sampling corneal reflection and pupil position at a rate of 1000 Hz. Viewing was binocular but only the dominant eye was tracked. Participants first completed the eye tracker's standard calibration procedure. Between each stimulus presentation a drift correction was performed. The distance between eyes and screen varied between 50 and 70 cm depending on each participant's natural posture. To familiarize participants with the experiment, they were first exposed to six randomized trials from the training set after which they had to complete 24 randomized trials from a different set. Rush-Hour problems differed with respect to the arrangement of vehicles. In particular, problems were manipulated with respect to the factor *has cluster* using a repeated measure design. A distracting cluster was present in 13 configurations.

We designed the experiment to being separable into two distinct stages: *planning phase* and *execution phase*. Planning takes place first, as soon as the participant receives the visual input of the board. Participants play the game through mentally and identify a solution: no interaction with the board is possible; they must therefore imagine all of the planning. An interactive experiment would give feedback to the participant on the validity of his moves and allow him to plan by trial and error. We want to avoid that to capture all the possible planning errors that a person can make.

Participants were unaware of the optimal solution's length, so they had no feedback that could tell them if the solution they found was optimal or not. This circumstance offered them the choice of keeping together or separating the clusters and, in turn, allowed us to determine their preferred strategy. After finishing planning, participant signaled this by a button press. This action represents the end of the first and beginning of the second phase.

During the execution phase participants were requested to



board	states	leaves	length	correct (%)	optimal (%)
left	457	214	6	83	83
right	432	235	6	77	0

Figure 4: Two configurations that have similar algorithmic properties but show a great difference in complexity

convey the planned solution by clicking on the individual cars in the image. The solution is recorded as an ordered list of color names that represent the order in which the vehicles have been moved. We cannot exclude that participants used the execution time to plan further, but it is unlikely given the short time recorded between clicks.

Results and discussion

Behavioral data

Table 1 reveals that the average correctness of the data collected from the execution phase is very similar for both types of configurations. In configurations lacking a cluster almost every correct solution is the optimal solution, while the optimality drops almost two thirds when the board presents one distracting cluster. Among the “difficult” problems the optimality drops considerably but the overall correctness remains largely constant.

Table 1: Solution quality by configuration type (in %).

	Solution	
Cluster	Correct	Optimal
Yes	85.4	31.3
No	80.6	77.0

We present an illustrative example in Fig. 4, where the two types of configurations show similar algorithmic properties but a huge gap in performance when played by human players: The rightmost configuration was optimally solved by the 83% of the participants, while the leftmost configuration was not optimally solved by any participant. Another interesting property is that in both configurations participants find a correct, suboptimal solution around 80% of the time. This finding indicates that participants have difficulties finding the optimal solution in the second board, while the difficulty of solving the game remains constant. The only difference that we could find between the two boards in Fig. 4 is the presence of a cluster in the right one. To test our assumptions that a cluster constitutes a Gestalt entity that effects solution optimality, boards were classified with respect to the property *has cluster*. The effect of the gestalt type of *has cluster* on the correctness and on the optimality of the solution, the latter only within the correct responses, was then tested in logistic mixed-effects models. As optimal implies correct, statistics on optimality have been computed on the subset of correct solutions. We included by-participant and by-item random intercepts to account for inter-individual differences among participants and items.

Table 2: Logistic mixed-effects model results for solution optimality and correctness.

Optimality

	b	*SE b*	*z*	p_z
(Intercept)	3.57	0.47	7.58	.00
Gestalt L1	-3.57	0.63	-5.66	.00
Gestalt L2	-4.69	0.60	-7.87	.00

Correctness

	b	*SE b*	*z*	p_z
(Intercept)	2.06	0.42	4.92	.00
Gestalt L1	0.59	0.53	1.12	.26
Gestalt L2	0.38	0.45	0.84	.40

The gestalt type of clusters had a significant effect on the optimality of the solution with respect to correctly solved problems, $\chi^2(2)=37.54$, $p < .001$. Table 2 (top) shows the respective results of a logistic mixed-effects model fitted to solution optimality indicating that giving an optimal response was less likely for problems that contained any type of cluster compared to those that contained no distracting cluster. Gestalt type, however, had no significant effect on correctness, $\chi^2(2)=1.38$, $p = .50$ (cf. the bottom of Table 2). This result was expected as the average correctness rate remains constant across all configurations.

Furthermore Type L2 had a stronger effect than Type L1. This effect was also expected as Gestalt Type L2 clusters show a higher degree of symmetry as both ends are aligned, therefore they form a stronger gestalt meta-object.

In sum, the results indicate that apart from algorithmic properties, such as branching factor or number of possible solutions, spatial properties also have a strong effect upon the performance, i.e. the optimality of a solution. In particular, we investigated the effect of the relative positioning of vehicles, i.e. the grouping principle of ‘proximity’ in classical Gestalt theory. Our hypothesis is that embodied processes create a bias towards the suboptimal solution whenever a cluster, a group of vehicles with specific reciprocal spatial relations, is present.

The first hypothesis is based on Gestalt theory: People tend to perceive certain clusters of objects as single entities, i.e. meta-objects that are then used in the planning phase. In the execution phase, however, the single components are

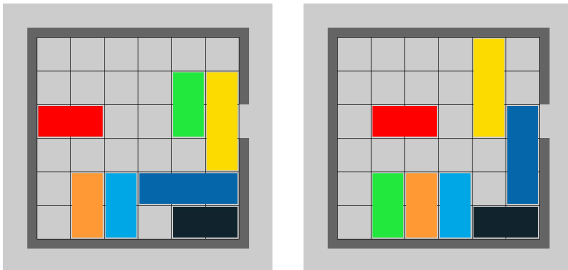


Figure 5: The cluster is first created in the planning phase and then separated into its components in the execution phase

perceived again as a result of actually moving these single components, i.e. vehicles in Rush Hour, to execute the just developed plan. Another possible explanation is the use of what we called *longest move strategy*: people avoid calculating exactly how long the vehicles are and what minimum distance they need to be moved to free the blocked car. Instead, they try to move them as far as possible; in this process each component of the cluster is seen as a separate entity but is an obstacle to the goal of sliding the blocked car as far as it can go.

It is well known that humans tend to solve computationally complex problems by chunking information in order to simplify the problem's representation (Ellis & Siegler, 1994; Kotovsky, Hayes, & Simon, 1985; Ohlsson, 1992).

Both explanations seem to be cognitively plausible because both remove some amount of complexity from the game. The Gestalt strategy reduces the number of vehicles that participants must consider to plan the solution by grouping them as meta-objects that are treated as single entities. On the other hand, the longest move strategy reduces the complexity due to distance calculations; participants do not need to check prior to every move if the space in the direction in which the vehicle has to be moved is enough to unblock the previous one in the move sequence. Participants can assume that, once a vehicle reaches the border in the other half of the board, the previous one will be automatically unblocked. The two theories lead to the same result and are difficult to distinguish from one another while observing participants performing the experiment. We have found evidence to support both theories, so we cannot say conclusively which one is correct.

Not only are the components moved individually, but sometimes their movements are even separated by other cars' movements, such as in Fig. 5.

In this example 40% of the participants, 70% of who solved the game by moving the cluster, freed the way for the leftmost car. To do so they moved in order one element of the cluster, the goal car and then the second element of the cluster. The solution that we expected, namely moving the goal car first and then sliding the cluster upwards, was performed by only 17% of the participants.

The objection here is that if participants must split the cluster while still in the planning phase, at the time they should also realize that there is a better solution and change their strategies accordingly.

Given these results we cannot draw a definite conclusion about which strategy is preferred by the participants. Our conclusion is that the two strategies are concurrent and are chosen depending on some criteria. From our current data we could deduce that people often change their strategies during the experiment and very few stick to the same strategy for the entire game. Unfortunately we cannot make any hypothesis on why and when people change their strategies because in our data participants seem to change strategies casually, without any dependence on board or time.

Eye-tracking data

Eye movement behavior for optimal and suboptimal solutions. In order to understand the underlying mental processes and the effect of Gestalt properties, we analyzed the eye-tracking data that were recorded while participants were planning their moves. We found differences in the eye-tracking patterns depending on whether or not a board was solved optimally.

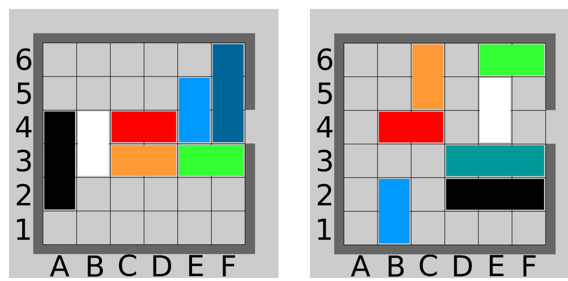


Figure 6: The importance of Gestalt in planning. The images have been post-processed for the sake of understandability

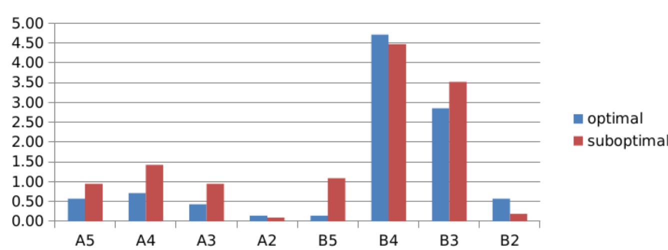


Figure 7: Average fixation per participant, coming from or going to the cluster, of the board presented in Fig. 6 left

The configuration in Fig. 6 (left) has one distracting cluster (white and black vehicles). We isolated all the fixations that fall inside the cluster. We found that participants' attention was concentrated on the white car, in particular on the cell B4. This finding suggests that participants recognize the white car as being crucial in planning.

To prove if the cluster is perceived as such, we took in consideration the immediately preceding and subsequent fixations. Players who perceive the cluster will try to move both elements in the same direction. If so they should look at the cells outside the cluster in the direction in which they want to move the vehicles, in our example upwards. As predicted (see Fig. 7) cells A5 and B5 attract the majority of preceding and subsequent fixations outside the cluster. The results support our hypothesis that the players want to move both elements of the cluster upwards, despite the fact that only the white car must be moved for an optimal solution. Moreover, if we discriminate between trials that have been solved optimally and those that have been solved suboptimally, we get more validation: in the optimal cases this effect is completely gone, as participants almost never looked at these cells. They instead fixated more often cell B2 which suggests the intention of moving the white car downwards therefore breaking the cluster. The black truck (A3 and A4) is fixated twice as often in case of a suboptimal solution, while the number of fixations on the white car (B3 and B4) does not vary much with solution quality.

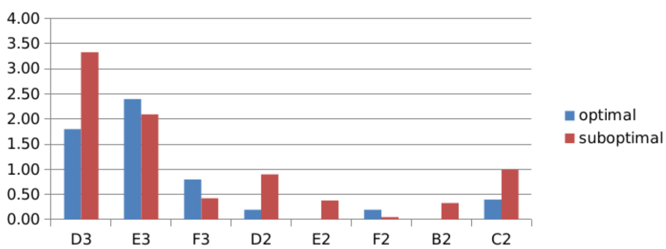


Figure 8: Average fixation per participant, coming from or going to the cluster, of the board presented in Fig. 6 right

Another interesting scenario is given by Fig. 6 (right). This configuration is special because both elements of the cluster have a distracting effect: The optimal solution does not require moving any of them. Nonetheless the solutions were biased towards the movement of the cluster. The fixation patterns show a difference between the data recorded for optimal and suboptimal performances: In optimal cases, only at the upper truck in the cluster was fixated, with the most fixations in cell E3. The behavioral data shown that this truck has not been moved, therefore it must have been considered only as blocking element for the white car and then excluded from the plan. In suboptimal cases, when the block has been moved, we found an increased interest for the black bottom truck (D2, E2 and F2). This time the cells with the most fixations are D3 and D2 (cf. Fig.8). This indicates an interest in the moving both trucks leftwards.

This is also supported by preceding and subsequent fixations. In optimal cases cell left of the trucks were never fixated, while in the suboptimal cases we recorded several fixations left of the black truck (C2 and B2).

Table 3: Logistic mixed-effects model results for the probability of fixating a field in the moving direction given by the optimal solution path.

Gestalt

	b	$SE\ b$	z	p_z
(Intercept)	-2.99	0.59	-5.10	.00
Gestalt L1	0.16	0.21	0.76	.45
Gestalt L2	0.70	0.19	3.75	.00
Cluster type (same dir)	0.97	0.40	2.40	.02
L1 × same.dir	0.30	0.39	0.77	.44
L2 × same.dir	-1.22	0.39	-3.11	.00

Effect of Gestalt on the perception of clusters. To prove the general validity of our findings, we analyzed fixations on clusters in all boards. Our analysis was restricted to clusters formed by only two vehicles. The focus was on those fixations that fall on elements of a cluster, and in particular on the landing position of their next fixation. Specifically, we looked at the probability of next fixations falling in the direction given by the optimal solution path.

Table 3 shows the results of a logistic mixed effects model that reveals that Gestalt level (see Fig. 2) has a reliable effect on the fixation behavior and interacts with Gestalt type.

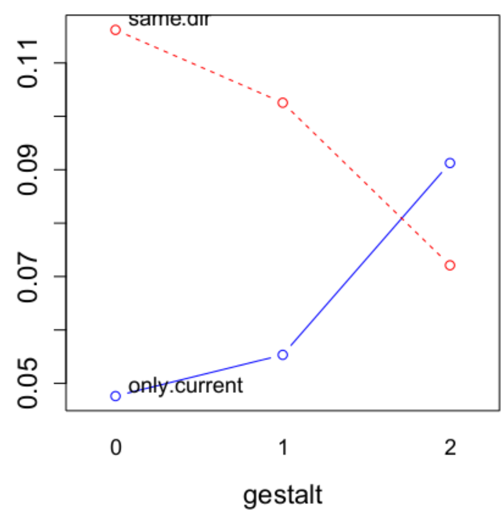


Figure 9: The probability of the next fixation landing on a field in moving direction according to the optimal solution path, as a function of Gestalt level and cluster type (same.dir: both cars must be moved in the same direction; only.current: only currently fixated car must be moved).

As illustrated in Fig. 9, the probability of fixating a field in moving direction increased with Gestalt level, but only for clusters in which the currently fixated vehicle had to be moved. This indicates that even when the second vehicle did not have to be moved, but was part of a cluster, it interfered with the perception of the currently fixated vehicle. By contrast, for clusters where both vehicles had to be moved in the same direction, a higher Gestalt level reduced the attention on fields in the moving direction. This result might indicate that less attention in moving direction is needed for the currently fixated vehicle when it shares the moving direction with the other vehicle of a well-formed cluster (Level 2).

This result shows that apart from structural properties of the problem, perceptual properties such as Gestalt of clusters can interfere with planning behavior.

Summary

In this article we studied how the spatial configuration of a board influences planning behavior. The novelty of our approach is that we did not study perception and planning separately, instead we looked for an interdependence between the two. We found a significant effect of the presence of clusters on the optimality of the solution: the presence of a cluster modifies the behavior of participants from an optimal to a suboptimal solution that involves moving the cluster even if only one vehicle must be moved. This suggests that visual perception, like Gestalt principles, can interfere with planning.

The experiment was divided in two separate stages: the planning and the execution phase. During the whole experiment participants were presented with a static image of the board. The absence of a visual feedback while solving the game might be seen as a limitation but is a feature instead: it excludes external memory and allows to capture all errors made during the planning phase to be captured. On the contrary, participants of an interactive experiment would have realized many incongruities of their plan during the execution phase. The experiment confirmed the substantial difference of performance with regards to optimality, strongly related to the presence of a cluster. This difference is not considered by any automatic planning algorithm, suggesting that human planning is influenced by factors that are not found in formal representation of the problem. On the other hand no substantial influence of clusters has been found on solution correctness or response time.

The lack of dependency for these parameters suggests that clusters need no cognitive effort to be processed, because the processing is done in a subsymbolic manner, and this points once again to Gestalt Theory. Two plausible heuristics have been introduced that could explain how these differences arise. The first heuristic is based on Gestalt principles: Clusters are seen as unique entities, following the Gestalt principles of proximity and continuity, and treated as atomic objects in planning. This reduces the cognitive load by decreasing the number of objects that need to be considered in planning. The second heuristic does not consider Gestalt effects but simplifies the planning phase by removing unnecessary spatial calculations while moving the cars: Vehicles are not moved the minimal sufficient distance but instead moved as far as possible. Both heuristics are compatible with the majority of the empirical results, but both are also disproved in a few situations, suggesting that these two strategies are complementary and concurrent and might both be present in human planning.

Future work will have to clarify the conditions in which one strategy is preferred over the other. To this end we are studying new boards that will concentrate on single interesting configurations and try to give more information about their complexity. Also a more thorough investigation of the eye-tracking data is needed to improve our understanding of how participants interpret the board representation and take advantage of it while planning.

Acknowledgements

This research was supported by the DFG (German National Research Foundation) in the Transregional Collaborative Research Center, SFB/TR 8 within project R8-[CSPACE]. Many thanks to Matthias Frorath and Corinna Blum for their assistance in data acquisition and preparation.

References

Ellis, S., & Siegler, R. S. (1994). Development of Problem Solving. In R. J. Sternberg (Ed.), (2nd ed., pp. 333–367). San Diego: Academic Press.

Flake, G. W., & Baum, E. B. (2002). Rush Hour is PSPACE-complete, or "Why you should generously tip parking lot attendants". *Theor. Comput. Sci.*, *270*(1-2), 895–911.

Kaller, C., Rahm, B., Köstering, L., & Unterrainer, J. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural brain research*, *216*.

Kaller, C., Unterrainer, J., Rahm, B., & Halsband, U. (2004). The impact of problem structure on planning: Insights from the Tower of London task. *Cognitive Brain Research*, *20*(3), 462–472.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace.

Köhler, W. (1959). Gestalt psychology today. *American Psychologist*, *14*(12), 727–734.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*, 248–294.

Morris, R., & Ward, G. (2005). *The cognitive psychology of planning* (K. J. Gilhooly, Ed.). New York: Psychology Press.

Ohlsson, S. (1992). Information processing explanations of insight and related phenomena. *Advances in the Psychology of Thinking*, *Vol.1*, 1-44.

Ragni, M., Steffenhagen, F., & Fangmeier, T. (2011). A Structural Complexity Measure for Predicting Human Planning Performance. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society.* Austin, TX.

Wertheimer, M. (1938). Laws of organization in perceptual forms. *A source book of Gestalt psychology*, 71–88.