A NETWORK ANALYTIC APPROACH FOR EXPLORING THE COMPLEXITY OF *RUSH HOUR*

MAREIKE BOCKHOLT*

1 INTRODUCTION

It is well known that problem solving capabilities of humans and computers differ in several aspects: While human can make use of their experiences, creativity, and some kind of intuition, computers must rely on the given data and algorithms in which not all real world constraints may be implemented. On the other hand, in processing and storing a big amount of information and dependencies, computers clearly outperform humans. Hence, it might be a promising approach to combine the structural advantages of human and artificial problem solving abilities in order to construct human-computer cooperative and interactive systems [1]. With the purpose of dividing subtasks between human and computer agents, it is necessary to better understand why some subtasks may be a challenge to solve for both. In computer sciences, complexity theory has been providing a broad range of results about problems' difficulty for being solved by algorithms. However, in cognitive sciences, there are only a few approaches to systematically analyze a problem's complexity for humans to solve it ([6], [5], [4]).

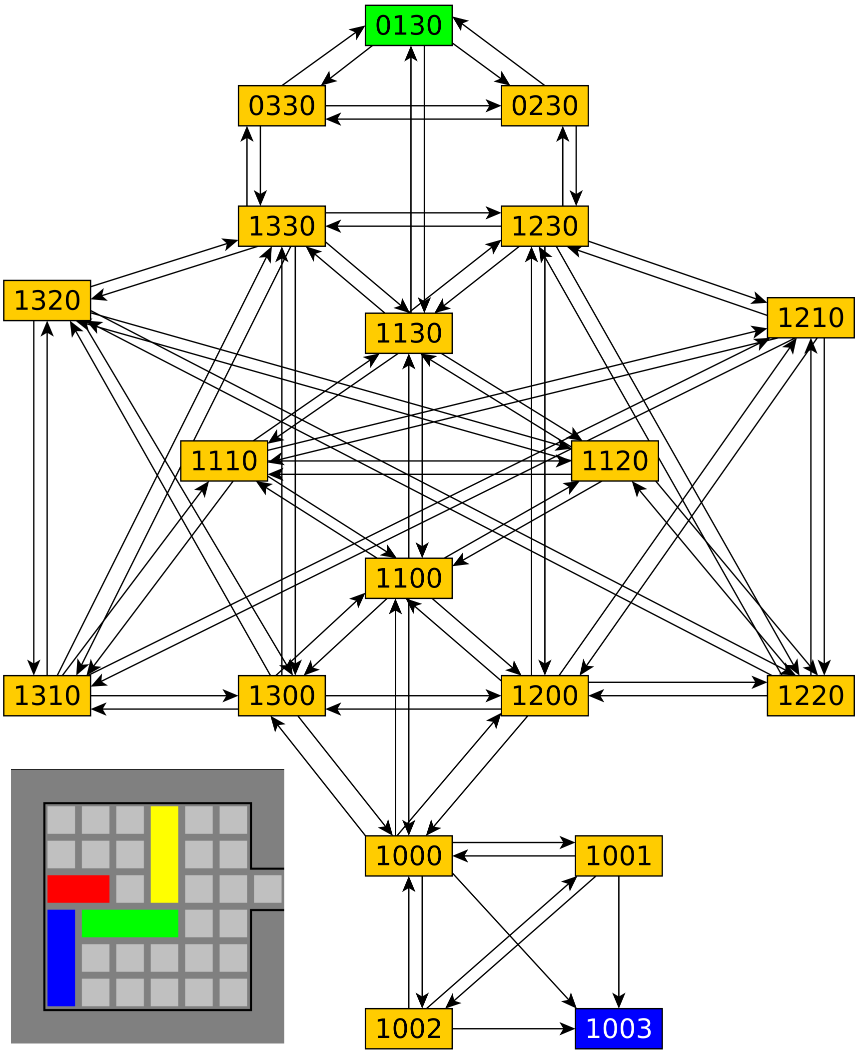
The present work uses the one-player game *Rush Hour* to exemplarily investigate the difficulty of a complex task for a human solver. But instead of only considering the game's intrinsic properties, we observe that a game's problem space – all from a starting configuration reachable game states with their transitions into each other – is a network (cf. figure 1 or 7). We expect a problem space's structure to reflect the complexity of its corresponding game. In this work, we introduce measures to capture a problem space's structure in order to investigate whether the structure and the corresponding game's difficulty correlate. Furthermore, we describe the findings of a conducted experiment in which the participants played some of the studied games which were selected due to their complexity measures. The results' analysis reveal essential flaws in human problem solving abilities which could be compensated by a computer-aided system.

2 APPROACH

Our research focus lies on the one-player board game *Rush Hour*, a sliding block puzzle game which takes place on a grid of 6×6 cells, representing a parking lot, with one exit (cf. figure 1). Cars of width 1 and length 2 respective 3 cells are placed on the board vertically or horizontally and can be moved forwards or backwards as long as the for the movement needed cells are not occupied by any other car. Cars cannot move sideways or rotate, and are not allowed to change their row or column, respectively. Given a

*mareike.bockholt@cs.uni-kl.de
Department of Computer Science
University of Kaiserslautern

Figure 1: A *Rush Hour* game instance with its corresponding problem space: the green state represents the shown configuration, the blue state corresponds to a final state, i.e. a state in which the red car can be moved through the exit, the transitions constitute all possible moves. The numbers in the nodes correspond to the positions of the cars in their row or column: the first/second/third/fourth digit corresponds to the red/green/blue/yellow car's position.



configuration of cars placed on the grid, the goal is to find a sequence of moves that allows a particular car (in figure 1 the red one) to be moved from the board through the designated exit.

It is obvious that the difficulty of finding a solution sequence is determined by the board configuration in the beginning. But is it possible to quantify the factors that contribute to the difficulty of a game? In order to answer this question, it might be useful to abstract from the explicit board configuration, but to consider the state space of it.

For every solvable board configuration, there is a unique state space which consists of the start configuration and all from there by allowed moves reachable board configurations. Two configurations (states) are linked if they can be transformed into each other by an allowed move (cf. figure 1). Finding a sequence of moves to solve the game can then be understood as finding a path through the state space from the start state to one of the solution states. The underlying idea of our research is that the difficulty of solving a game, meaning finding a path through the state space, should depend on the state space's structure. State spaces of different structures should yield games of different levels of difficulty. The following section deals with the question of how the structure of a state space can be quantified in order to compute the correlation between a game's complexity and its state space's structure.

2.1 *Measuring a network's structure*

In the following, several network measures are introduced which could capture a network's structure and be associated with the difficulty of solving the corresponding problem. In the following, the measures are introduced in an intuitive and informal way. The appendix contains a formal description (see section A.1). We propose the following measures:

SIZE OF STATE SPACE Since a game could turn out to be more difficult to solve if there are more possible states to explore, we introduce three different measures based on the size of the state space: number of reachable states (*nodes*), number of reachable states without the final states (*nodf*), and number of possible moves (*edges*).

LENGTH OF SOLUTION The number of needed moves to reach a final state should increase the difficulty of a game, therefore, we introduce the measure length of shortest path (*lsp*) which is the minimal number of moves from the start configuration to any of the final states. It can be observed that a state space can have different final states which have different distances to the start configuration – even if one takes the shortest possible path. To model this observation we introduce a measure based on the length of the solution path: the average solution length (*avlsp*) which is defined as average number of moves to a final configuration (if the shortest path is used).

NUMBER OF DECISION POSSIBILITIES In every configuration the player has several possibilities which move next to take. We hypothesize that a game should be harder to solve if there are more possibilities to consider in every move. Therefore, we introduce the following metrics: the average degree (*avdg*) is defined as the averaged number of decision possibilities, taken over all states except the final states. Since the state space might contain a huge number of states of which the most players only explore a small fraction, we approximate this fact by restricting the state space to a smaller one, namely the one which only contains states on shortest paths to a final state. Basic assumption is here that this will approximately be the part of the state space which most players will use for their solution. Counting the decision possibilities a player has in this restricted state space (the possibility of leaving the restricted state space included resp. excluded) yields, averaged over the number of states considered, the measure *avdgog* resp. *avdgop*. We observe that there are many states in which a great number of moves are possible, but most of them belong to optimal solution paths. Making a right decision should be harder if the ratio of good decisions to the number of all possibilities is small than if there are only good choices to make. Therefore, we propose the measure of branching complexity (*br*): we define the *branching complexity* of a single node as the number of bad choices (i. e.the number of possible moves which do not belong to the optimal solution path) divided by the total number of choices. The *branching complexity* for the network is then the average of the branching complexity of all nodes.

NUMBER OF OPTIMAL PATHS Furthermore, we hypothesize that the number of different optimal paths could influence the difficulty of a game. For this reason, consider the measure shortest paths (*sp*) which counts the number of possible optimal paths from the start configuration to a final one, and the measure shortest paths per final state (*sppf*) which is *sp* divided by the number of final states.

GAME PROPERTIES Up to now, only properties of the state space were considered such that the aforementioned measures could also be applied to any other board game for which the concept of a state space makes sense. Therefore, we also want to consider game specific measures:

- The simplest approaches only use the number of cars (*cars*) a configuration contains respectively the number of occupied cells on the board (*fields*), since handling more movable objects in finding a solution is supposed to be cognitively more challenging.
- On the other hand, having more cars on the board often means that the cars block each other such that there are effectively less objects to handle. For that reason, the average number of movable cars in every configuration is calculated and taken as measure *mc* (which is similar, but not the same as *avdg*). For the same reason as above, we also consider this measure on the restricted state space of optimal paths (*mcop*).

NOT INTUITIVE MOVES From research in human problem solving, it is well known which heuristics humans apply for solving a task. One of them is called hill-climbing [2]: the current situation is compared with the desired one, the operator which yields a more similar situation to the solution is chosen. In our game, the goal is to unblock the red car and move it forward to the exit. A human playing according to the hill-climbing method, will try to successively remove the blocking cars out of the way of the red car and successively move it towards the exit. Though, there are a lot of board configurations for whose solution it is necessary to move the red car backwards or to temporarily block the red car by another car. Because this kind of moves contradicts the hill-climbing method, we call these moves *counterintuitive moves* and suppose that a larger number of counterintuitive moves needed in a solution should increase the difficulty of the game. Therefore, we define the number of counterintuitive moves as measure, weighted by the factor in how many solution paths this counterintuitive moves appears (*cm*). Since longer solution paths are expected to contain more counterintuitive moves, but the length of the solution path is already represented in the measure *lsp*, we normalize *cm* by *lsp* and get the measure *cmpl*.

An overview of the introduced measures and their range of values can be found in table 1.

MEASURE	DEFINITION / EXPLANATION	RANGE
nodes	number of states in problem space	$\mathbb{N}$
nodf	number of states in problem space without final states	$\mathbb{N}$
edges	number of transitions in problem space	$\mathbb{N}$
lsp	length of optimal solution	$\mathbb{N}$
avlsp	average length of solution	$\mathbb{R}_{\geqslant 0}$
avdg	average number of decision possibilities	$\mathbb{R}_{\geqslant 0}$
avdgop	average number of decision possibilities on optimal paths (and to stay on optimal paths)	$\mathbb{R}_{\geqslant 0}$
avdgog	average number of decision possibilities on optimal paths	$\mathbb{R}_{\geqslant 0}$
br	average fraction of transitions that lead away from optimal paths	$[0, 1]$
sp	number of shortest paths from the start configuration to a final state	$\mathbb{N}$
sppf	number of shortest paths per final state	$\mathbb{R}_{\geqslant 0}$
cars	number of cars	$\{0, \ldots, 18\}$
fields	number of occupied cells	$\{0, \ldots, 36\}$
mc	average number of movable cars	$\mathbb{R}_{\geqslant 0}$
mcop	average number of movable cars on optimal paths	$\mathbb{R}_{\geqslant 0}$
cm	weighted number of counterintuitive moves	$\mathbb{R}_{\geqslant 0}$
cmpl	weighted number of counterintuitive moves in relation to length of solution	$\mathbb{R}_{\geqslant 0}$

Table 1: A summary of the used complexity measures.

2.2 *Data*

We used the level card packs that are included in *Thinkfun*'s *Rush Hour* game. There are three standard level card packs (1 (regular edition), *deluxe* and *junior* edition) as well as three additional level card packs (2,3, and 4). Each card pack contains 40 (*deluxe*:60) different start configurations whereas a difficulty estimation is assigned to each start configuration by *Thinkfun* (beginner, intermediate, advanced, expert, and grand master). In the following, the configurations from every card pack are used, except of cards from the *Junior* edition. Configurations in which two red cars instead of one need to be removed from the board are excluded from further analysis as well as identical configurations from different packs are used only once. Table 2 shows how many level configurations from which card set and in which difficulty level were available and used for the following analysis.

	B	I	A	E	G	SUM
Junior	10	10	10	10	0	40
Standard	10	10	10	10	0	40
Deluxe	10	3	3	3	8	27
Set 2	0	10	9	6	7	32
Set 3	0	10	10	9	9	38
Set 4	0	10	10	8	8	36
SUM	20	43	42	36	32	173

Table 2: Number of used start configurations of each card pack *Thinkfun* provided, ordered by its difficulty rating: beginner (B), intermediate (I), advanced (A), expert (E), grand master (G).

3 ANALYSIS AND RESULTS

The above introduced measures were computed for all available level cards and are analyzed in order to discover a correlation between the value of the measure and the difficulty rating given by *Thinkfun*. The results are visualized in a box plot diagram shown in figure 2 in which each measure's values are plotted, ordered by difficulty level (beginner, intermediate, advanced, expert, and grand master). Each box contains 50 % of the data, the black bar in the box represents the median of the data. A box' whiskers show the minimum respective maximum data point which is no more than 1,5 times the interquartile range from the box. Values outside of this range are shown as outliers as single points. If a measure perfectly correlated with the difficulty rating, the boxes would be flat without any whiskers or outliers and would lie on an ascending line.

Considering figure 2, it can be clearly seen that this perfect correlation does not occur for any of the proposed measures. Though, one can observe a correlation between the difficulty and the measures *lsp* and *avlsp* which are both related with the length of the solution path. This observation is in agreement with the findings of Ragni et. al. [6]. Though, it is remarkable that even the solution path length based measures have a considerably large range and even outliers. Under the assumption that the difficulty rating is correct, there must be easy games with a long solution path as well as hard games with an unusually short solution path. Therefore, there must be further factors which contribute to the difficulty of a game.

All other measures do not allow any clear conclusions. Table 3 in the appendix shows the mean values and standard deviation of all measure values, ordered by the game's difficulty rating. Furthermore, one column

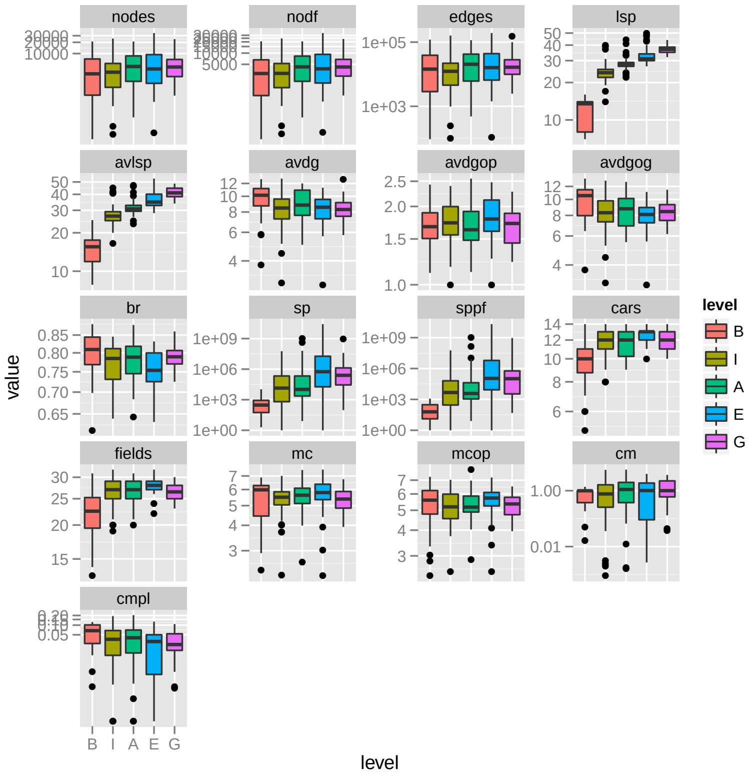


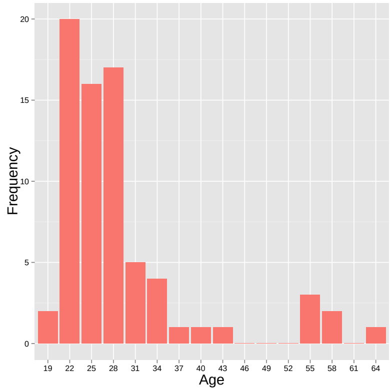
Figure 2: Values of the complexity measures, visualized with a box plot, separated by difficulty classification (beginner (B), intermediate (I), advanced (A), expert (E), grand master (G)). The scale is logarithmic. Games from the junior edition are excluded from analysis.

displays the correlation of the measure's value with the difficulty rating (Pearson's correlation coefficient). As it can be already seen in figure 2, only the solution length based measures *lsp* and *avlsp* show a correlation with the difficulty rating of the game. All other measures seem to be, considered as single predictors, unrelated to the difficulty of the game. But although the length of solution path turns out to be the single measure which shows a strong correlation to the difficulty rating of the game, it might be worth to look at the games which do not fulfill this relationship as there must be other factors contributing to the true complexity of these games.

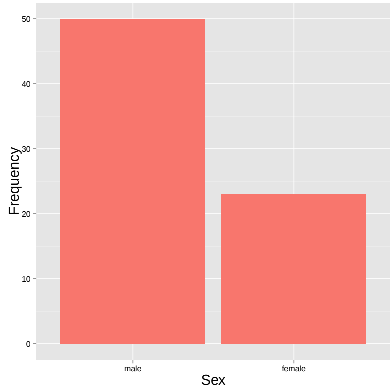
In order to find out which other factors there might be, out of the 173 games of interest, 24 games are selected due to their complexity measures and their difficulty is tested in an experiment.

4 EXPERIMENT AND RESULTS

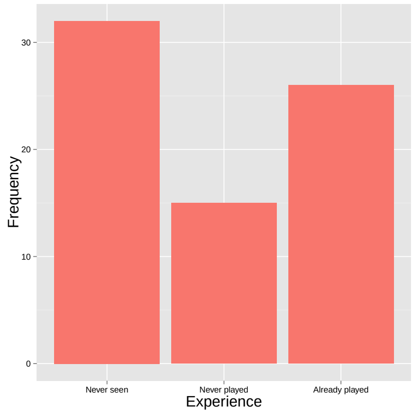
An experiment was conducted for which 24 games were selected, a part of them having particularly remarkable values in some measures, a part of them having measure values close to the mean value of their difficulty category. The participants were asked to play at least six of the 24 games of increasing difficulty. It was made sure that each participant played each game at most once. They had the possibility to quit a game and continue with the next game, but they did not have the chance to resume an already started game. The participants were asked to find a short solution, i. e. a solution with a minimum number of moves. When a participant solved a game, he/she was asked to rate the game by its difficulty. For the following analysis, 97 data sets, each containing 6 completed games, were used.



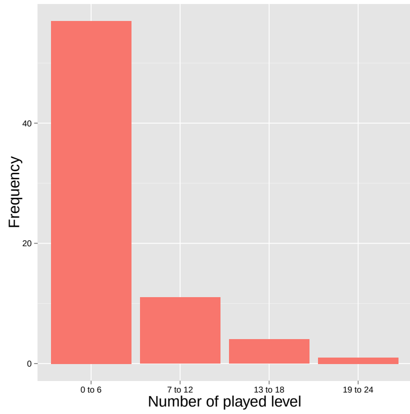
(a) A histogram that shows the age distribution of the experiment's participants.



(b) A histogram that shows the gender distribution of the participants.



(c) A histogram that shows the participants' prior experiences concerning the game *Rush Hour*.



(d) A histogram that shows how many games the participants played during the experiment.

Figure 3: An overview of the information we have about the experiment's participants.

4.1 *Selection of the level*

Table 4 in the appendix shows the measure values of the for the experiment selected games. The selection contains games whose measure values are quite close to the measures' mean values of the category, but also games which have measure values which are exceptions in their category. The measure which was the reason to select the game, is highlighted in red. As it can be seen in table 4, there is a beginners game with an exceptionally large state space (card pack:*deluxe*; card:02), an intermediate game with a remarkable short (card pack:*standard*; card:15) and one with a remarkable long solution path (card pack:*standard*; card:19). Among the advanced games, a game with an relatively low average degree (card pack:*standard*; card:26) was chosen as well as a game which is unusual in several of the measures (card pack:*standard*; card:22). The selected expert games contain a game with a low average degree, only one optimal solution and a high number of counterintuitive moves (card pack:2; card:28), and a game with an exceptionally large state space (card pack:*standard*; card:32). As grand master games, there was, among two others, chosen a game with a high number of counterintuitive moves (card pack:*deluxe*; card:55).

4.2 *Players*

There were 74 players participating in the experiment, an overview of the information about them can be found in figure 3. It can be seen that most of the players are in their twenties, about two third of the players are male, and the majority of the players did not know the game *Rush Hour* before.

4.3 *Solved and skipped level*

Each game chosen for the experiment was played by at least 20 participants. We first consider how many of the players who started a game were able to finish or even solve it optimally. For this purpose, consider figure 4 which shows this relationship. For each game, it is plotted which fraction out of all players who started this game were able to solve the game or even find the optimal solution.

The games are ordered by increasing difficulty such that the by *Thinkfun* as easy rated games are on the left and the games rated as hard are on the right. It is remarkable that even among the games which are rated as easy, there is none which was solved optimally by all of the players. There are only three games which were solved optimally by more than half of the players. Among the games rated as intermediate or harder, there are only very few players who were able to solve the game within a minimum number of moves. The game *deluxe 02B* was chosen because of his extraordinarily large state space, but it was optimally solved by a remarkably high number of players, it was not even skipped once. This gives a hint that a large state space does not significantly determine the complexity of the game.

Among the beginner games, the game which was skipped the most often and could be optimally solved the least often, is the game *1 10B* which was chosen because of his low average degree, meaning the low average number of move possibilities. The relatively high failure rate for this game can not be explained by the length of the solution path because the game *1 03B* with a similarly long solution path clearly shows a better success rate. At the same time, the expert game *2 28E* which was also chosen because of its low average degree has a high failure quote as well.

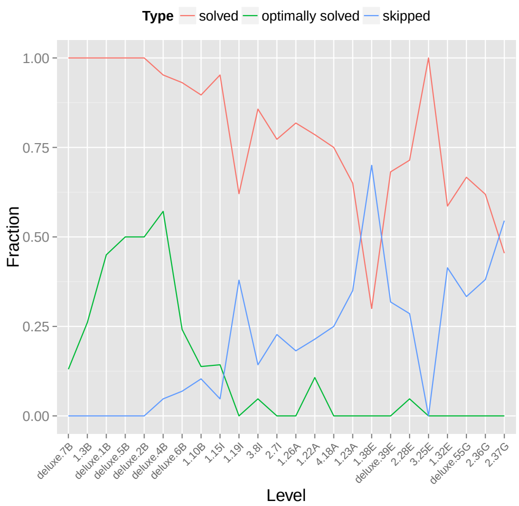
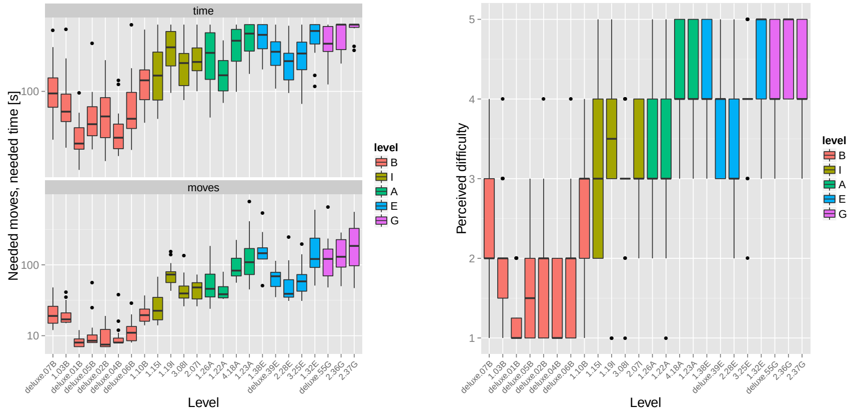


Figure 4: Fraction of solved, skipped and optimally solved games.

It strikes that the game with the highest skipping rate is not a game which had been classified as *grand master*, but as *expert*. But this can be explained by the unusually high number of needed moves for solving the game *1 38E*, it is the highest among all chosen games. Therefore, it is not surprising that many players did not finish the game.

In the following, the time and moves the participants needed to complete a game, are further analyzed. The participants did not play the games in a controlled environment, but on their own computers, since it was a browser-based game which could be played on-line. Therefore, we do not have certainty whether the participants were actively playing after the game was started or whether they might be distracted. In order to avoid too skewed results, all moves which took more than ten minutes are set to ten minutes and all further analysis will be done with these modified times. In addition to the players needed time and moves, the players' difficulty rating can be used for analysis. These information are displayed in figure 5: the horizontal



(a) An overview of how many moves and how many time the participants needed to solve the games. The color encodes the difficulty classification of the games.

(b) An overview of how difficult the games were rated by the players after they have finished the game.

Figure 5: An overview of the participants' solutions of the games. Only solved games are considered in both plots.

axis lists all in the experiment played games, ordered by increasing difficulty rating, the vertical axis indicates the number of moves and the time the participants needed to solve the game (in figure 5a) respectively the difficulty rating the participants gave the game (in figure 5b).

The rough relationship between the difficulty classification by *Thinkfun* and the time respectively the number of moves the participants needed can be observed here as well. Though, it is striking that there are games which do not fit in this pattern: Among all beginner games, the game *1 10B* seems to be the hardest one, since the participants needed the most moves and the longest time to solve it. This can not be explained by the length of its optimal solution path because the game *1 03B* requires a longer solution path, but was solved quicker by the participants and rated less difficult.

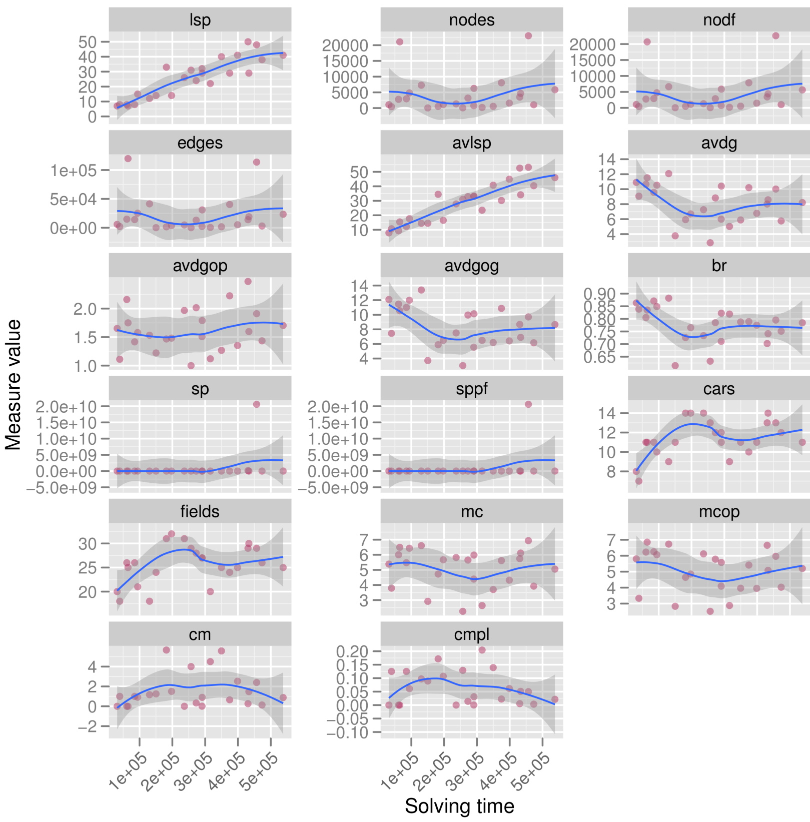
Furthermore, it is remarkable that the game *1 15I* shows a large variation in its difficulty rating although the variation in time and moves is not noticeable. Having a look at the game's measures, it can be seen that this game has a remarkably short optimal solution path, but it was rated as difficult by several players.

4.4 *Correlation to the measures*

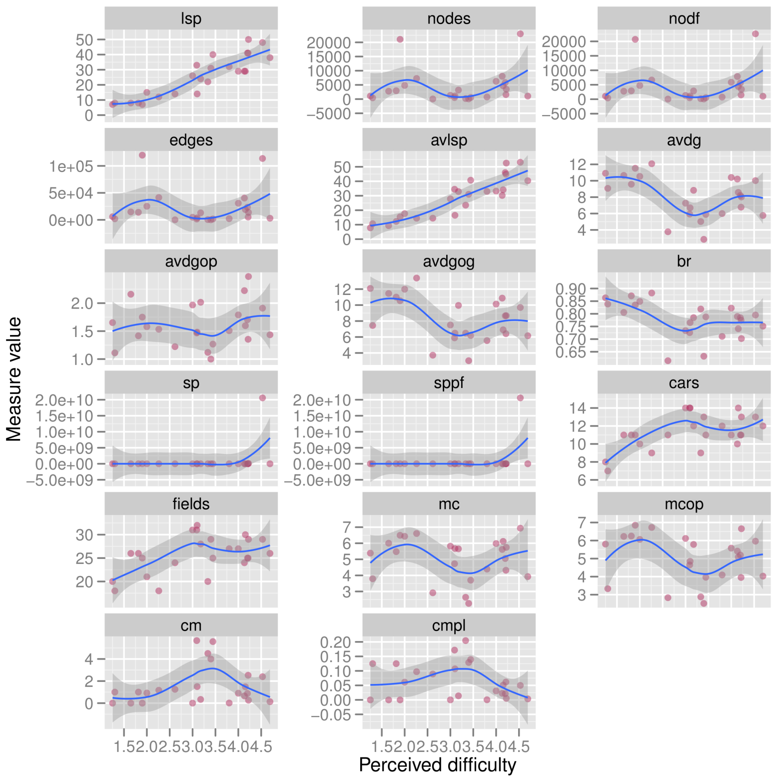
In the former sections, we found that our proposed network based complexity measures do not correlate with the difficulty classification of *Thinkfun*, except of the solution length based measures. Furthermore, first results of the conducted experiment were described. Having several different methods at hand to measure the difficulty of a game (classification by *Thinkfun*, number of moves or time needed by the participants, or participants' rating), it can be analyzed if the in section 2.1 introduced measures correlate with any of these.

Figure 6 visualizes this relation: in both figures, each diagram represents one of the introduced measures (cf. section 2.1), each point in the diagrams represents one of the 24 chosen games. The average needed time respectively the average difficulty rating given by the participants for every game is plotted against the game's measure value. It can be seen that the former result that the length of the solution correlates with the difficulty of the game, is confirmed here. Solving time as well as rated difficulty is correlated with the length of the optimal solution path.

In addition to that, none of the measures shows a clear correlation to the complexity of the games. Though, the degree based measures tend to have a



(a) Complexity measures plotted versus the solution time.



(b) Complexity measures plotted versus the participants' difficulty rating.

Figure 6: Relation between the proposed complexity measures and the participants' needed time respectively their difficulty rating.

slight negative correlation with needed time and difficulty rating: the less move possibilities there are, the more difficult the game is perceived and the more time is needed to solve the game.

A further observation involves the measures *cars* and *fields* whose value seems to contribute to the difficulty of the game up to a certain degree from which on the difficulty is independent from them. This finding can be explained intuitively: having more cars on the board may increase the difficulty at first because more objects need to be considered in order to find a solution (see also [4]). But from a certain number of cars on, the pure number of cars is not the main factor anymore which determines the difficulty, but their positions, if they block each other, etc.

4.5 *Getting lost in the state space*

During the experiment, several participants wished for the possibility of restarting a game which gives a hint that the navigation of the participants through the problem space might be worth to have a look at. In figure 7, the problem space of game *1 19I* is shown and how the participants navigated through it. In this and also in the visualizations for the other games, it is clearly recognizable that the participants preferred to take almost the same routes through the problem space which is not necessarily the shortest one. This fact supports the assumption that the players are guided by the same heuristics in their solution strategy, known from human problem solving research.

The question why it is more difficult for some games than for others to find a solution way through the problem space, is still not answered. In order to approach this question, the navigation of the participants through the problem spaces is examined closer. The assumption is that participants lose their way while finding a solution which could give them the impression that the game is harder.

For this purpose, it is essential to find a quantification for *getting lost* or *losing one's way*. The first naive approach is based on the simple idea: if a player struggles with finding a way to a final state, he or she will surely need more moves than necessary. Therefore, we consider how many moves the players needed in relation to the number of necessary moves. A visualization of this analysis is shown in figure 8.

It is interesting that the relative path length (number of moves done by the player divided by the number of moves in the optimal solution) correlates well with the difficulty rating by the players as well with the difficulty classification by *Thinkfun* – at least for the easier games. The games classified as harder (advanced, expert, and grand master) do not show such a clear connection neither to th difficulty rating of the participants nor to the relative path length. Furthermore, it is striking that range of the relative path length grows with increasing difficulty classification by *Thinkfun*.

Considering figure 8b, one can observe that the number of unnecessary moves the players did while playing, is directly related to their own difficulty estimation of the game: the more unnecessary moves they did – the more they got lost in the state space –, the more difficult the game is perceived. The difficulty classification by *Thinkfun* does not seem to contribute much, since the *Thinkfun* classifications are spread over all difficulty ratings of the participants in figure 8b. The perceived difficulty of a game seems to depend more on the player's disorientation than on the difficulty classification (whereas the classification might influence the extent of disorientation, but does not predict it perfectly).

Even the simple approach described above for capturing the concept of getting lost in the state space, shows a clear correlation to the perceived difficulty of the games. But since it is already known that the perceived difficulty is dependent from the length of the solution path, the observed effect might be a result of this dependence. In order to exclude this possibility,

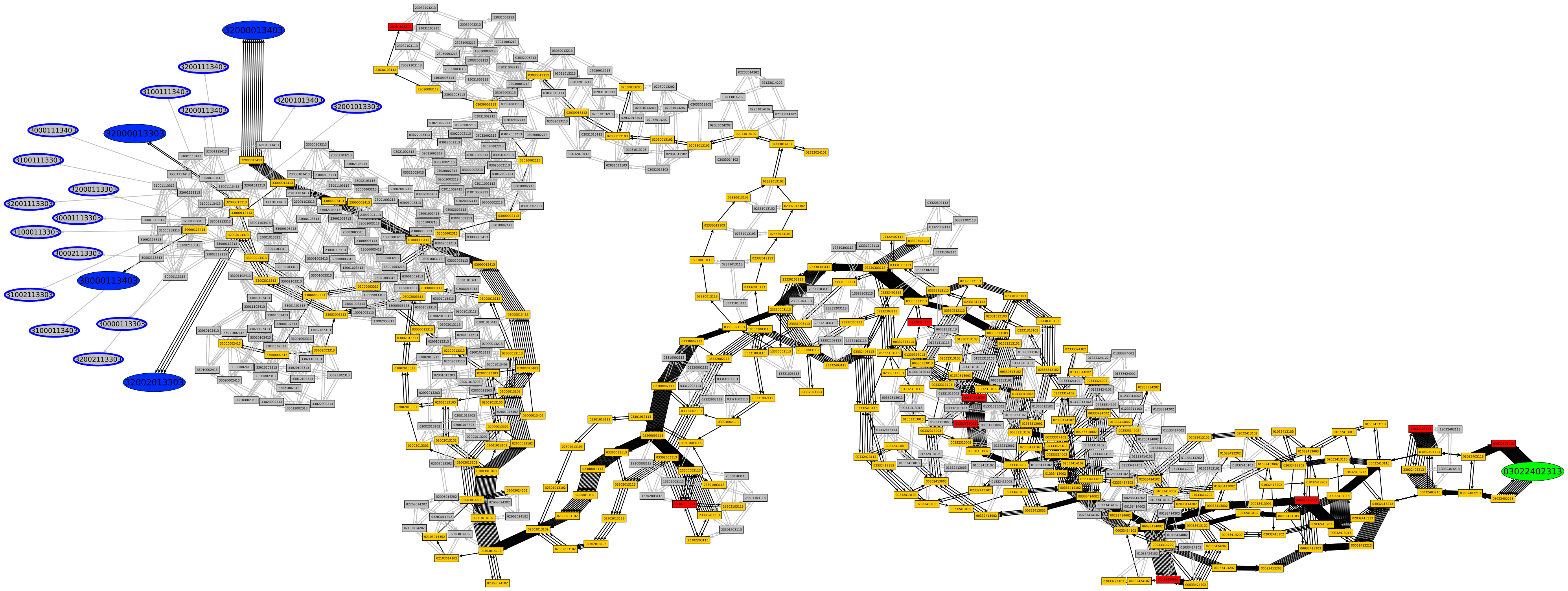
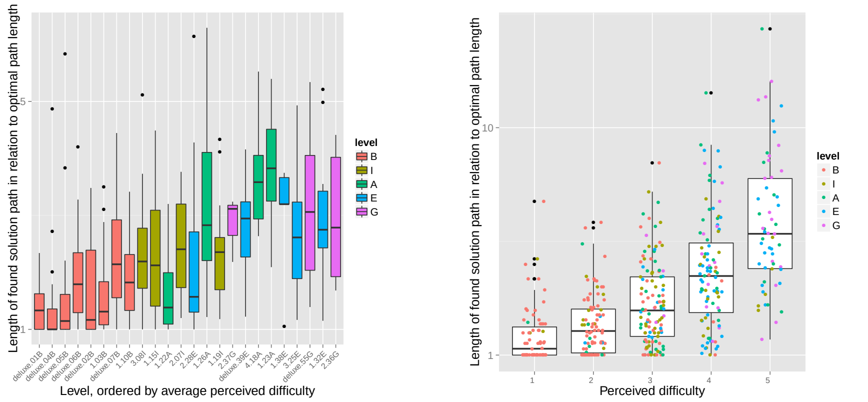


Figure 7: A visualization of the problem space of game *1 19I* and how the participants navigated through it while playing. Start node is green, visited nodes are yellow, not visited nodes are gray. Reached final nodes are blue, not reached final nodes are blue-rimmed. Nodes in which at least one participant quited the game are red. The number of edges shows how many participants took this transformation.



(a) On the horizontal axis, all played games are shown, ordered by their average difficulty rating by the participants. The vertical axis shows how many moves the participants needed to solve the level, in relation to the length of the optimal solution of the particular game. The color encodes the difficulty classification by *Thinkfun*.

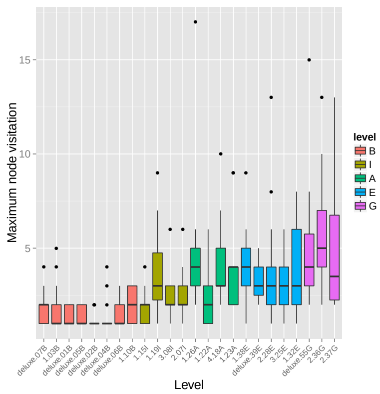
(b) The vertical axis also shows the ratio of needed moves to necessary moves, but aggregated by the participants difficulty ratings (1 *very easy* to 5 *very hard*). Each data point is one solved game of one player.

Figure 8: The relation of the relative path length and difficulty rating of the participants. All by participants solved games are contained in the plots. Note the logarithmic scale on the vertical axis.

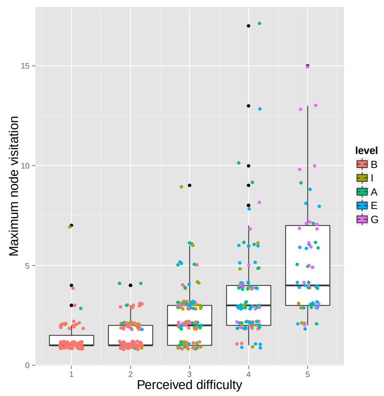
we consider two other approaches to formalize the concept of losing the orientation in the state space: the *average node visitation* and *maximum node visitation*. The underlying idea is as follows: if a player has difficulties to find a solution path or find a path to leave one part of the network and reach another, he or she might visit one state several times. Clearly, this does not cover all cases, since a player can be lost without visiting any state twice, but observing that the player visits states several times is a definite indication that he or she lost the way. Based on this thought, the *node visitation* for each node of a state space of a game for one player is defined as the number of times the player visited this node while playing. As a measure for losing orientation, the *average node visitation* and *maximum node visitation* are considered: the former being the mean value of all node visitations of all by the player visited nodes, the latter being the maximum value for one player and one game.

Figure 9 shows the *maximum* respectively *average node visitation* of all games and all players, ordered by game or ordered by difficulty rating of the players. At first view, one can observe that the figures – ignoring the scale – are similar to each other which can be explained by their similar underlying ideas. More importantly, both figures reveal a significant relation of the node visitations to the difficulty classification of the players. Therefore, the degree of how much a player loses orientation is a good indicator for how difficult he or she will perceive the game. This leads to the conclusion that a problem's complexity does not only depend on objective properties, but there is also a high correlation with the individual performance. Though, the question why players lose orientation is still open.

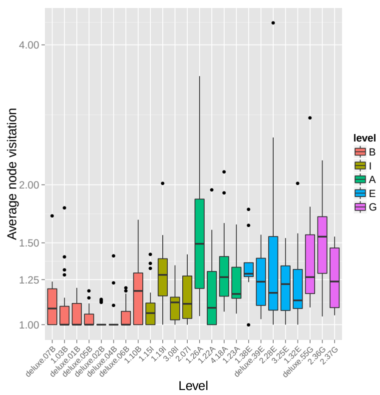
However, the values of the maximum node visitation take on surprisingly high values indicating that getting lost in a huge problem space is a general issue in human problem solving. But this could, though, be easily avoided by computer support, since recognizing a repeating configuration can be done algorithmically without need of completely solving or even knowing the problem. Thus, this analysis of a problem which may seem artificially constructed leads to the suggestion of the following human-computer cooperative system: the human can make use of intuition, creativity, and heuristics to solve the problem with a problem space which may be too large for a



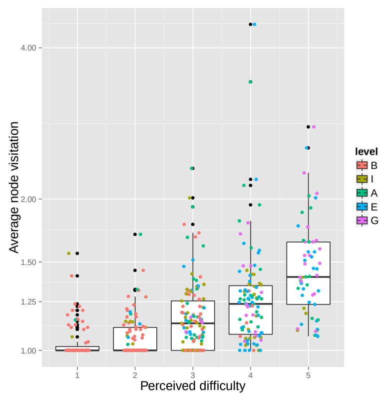
(a) *Maximum node visitation* plotted versus the single games. The box plots contain all *maximum node visitations* of all participants who played this game.



(b) *Maximum node visitation* plotted versus the difficulty rating of the players. The data points were arbitrarily scattered in horizontal direction.

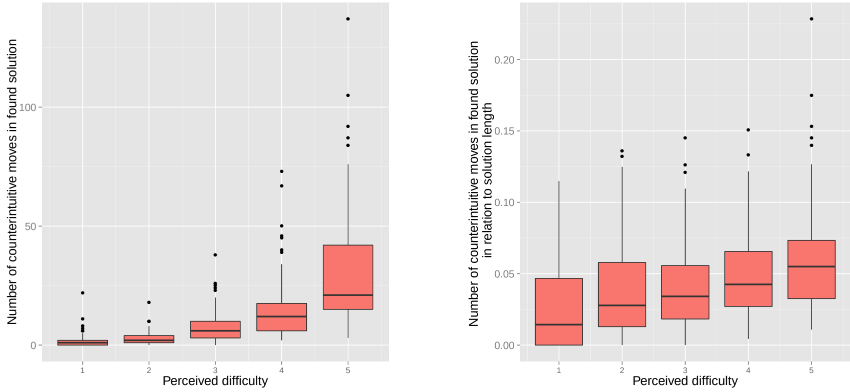


(c) *Average node visitation* plotted versus the single games. A box contains all *average node visitations* of all participants who played this game. Note the logarithmic scale on the vertical axis.



(d) *Average node visitation* versus difficulty rating of the players. Data points were arbitrarily scattered in horizontal direction. Note the logarithmic scale on the vertical axis.

Figure 9: An overview of the *node visitations* in all games.



(a) The relation of the player's difficulty rating and how many counterintuitive moves he or she used in the solution path. The difficulty rating range from 1 (very easy) to 5 (very hard).

(b) The relation of the player's difficulty rating and how many counterintuitive he or she used in the solution path, normalized by the total number of made moves.

Figure 10: Relation of difficulty rating and number of counterintuitive moves.

purely algorithmical solution, and the computer gives notice of repeating configurations, based on local computations, pointing the human to the right direction.

4.6 *Counterintuitive moves*

In section 2.1, the concept of counterintuitive moves was introduced. The experiment's results will be analyzed under this aspect in the following sections.

There are several aspects which makes the analysis of counterintuitive moves complicated. The assumption is that the proposed concept of counterintuitive moves are a kind of moves that players try to avoid. Therefore, players might take longer paths in order to avoid or to put them off for as long as possible. It might even be that players hesitate longer before they take a move when a counterintuitive move is due. The confirmation or rejection of these hypotheses poses some problems, though, which should be stated first.

On the one hand, players will plan several moves ahead, therefore, the time a player needs to take a move, is not necessarily related with the move directly following, but with moves that might come later. On the other hand, it might happen that players anticipate that a counterintuitive move will be the result of a particular sequence of moves, and they will not choose this sequence of moves, but prefer another in which the counterintuitive move is not contained. Finally, the fact that a counterintuitive move is possible does not mean that it makes sense to choose this move. For example, reversing a move in which the red car is moved ahead or unblocked – which needs to be done to solve the game – is a counterintuitive move in our sense.

We will therefore focus on the question whether the number of counterintuitive moves in a solution does have an effect on the difficulty of a game. Since the (weighted, but) absolute number of counterintuitive moves in the state space does not have any effect on the complexity of the games, as shown in sections 3 and 4.4, we consider the number of counterintuitive moves contained in the individual players' solutions in relation to the individual difficulty ratings. It is neglected whether the made counterintuitive moves are useful or not. The number of counterintuitive moves the players did in their solution paths (not: had to do) versus their difficulty estimation of the game is visualized in figure 10a. A strong correlation is observed: the more counterintuitive moves are contained in the individual player's solution, the more difficult the game is rated. However, the contribution of the absolute number of moves to the difficulty of the game is already shown, and a longer solution will contain more counterintuitive moves. Hence, the number of counterintuitive moves in the solution, normalized by the total number of moves in the solution is considered in figure 10b. A correlation to the game's difficulty can still be observed. So, players perceive games in which solutions they use more counterintuitive moves as more complex.

5 SUMMARY

In the described research, different approaches were used to identify factors which contribute to the complexity of the board game *Rush Hour*. A purely computational approach based on the problem space of the game revealed the significant correlation of the difficulty of the game and the length of its solution. Other network analytic measures did not show any significant dependencies on the difficulty classifications. Based on a conducted study involving 97 subjects, the finding that the length of the solution path strongly influences the difficulty of solving the game, is confirmed. Furthermore, several independent ideas to model disorientation in the problem space indicate that the individual player's performance has a greater influence on the perception of difficulty than assumed. In addition to that, the pure (and normalized) number of counterintuitive moves the players had in their solutions show a correlation to the players' difficulty rating.

A APPENDIX

A.1 *Formal description of the game and measures*

THE GAME

BOARD A *board* is defined as a set of cells: $B = \{(i,j) \in I \times I \,|\, I = \{0,\ldots,5\}\}$. A cell $b \in B$ can either be free or occupied.

CONFIGURATION Let $C = \{0,\ldots,n\}$ be the set of cars placed on the board. We define a *configuration* belonging to a board as

$$k = (C, pos_k, hv_k, cr_k, le_k)$$

with $pos_k : C \to I$, $hv_k : C \to \{0,1\}$, $cr_k : C \to I$ and $le_K : C \to \{2,3\}$. Interpreted in the graphical sense: C is the set of placed cars on the board. A car can be placed either horizontally or vertically, if car $i \in C$ is placed horizontally, $hv_k(i) = 1$ holds, otherwise $hv_k(i) = 0$. $le_k(i)$ specifies the length of car i. $cr_k(i)$ states the column number (for a vertically placed car) respectively the row number (for a horizontally placed car) in which the car is placed. $pos_k(i)$ indicates the minimal row index (for a vertically placed car) respectively the minimal column index (for a horizontally placed car) in which i occupies cells.

Consider configuration k in figure 1, for the green car, it holds: $pos_k = 1$, $cr_k = 3$, $hv_k = 1$ und $le_k = 3$.

VALIDITY Let $o : C \times B \to \{0,1\}$ be a mapping which, for a given car $c \in C$ and a given cell $b \in B$, returns 1 if b is occupied by c. We call a configuration *valid*, if the following holds:

i) $(\forall b \in B\ \forall i \in C : o(i,b) = 0) \vee$
$(o(j,b) = 1 \Rightarrow\ o(i,b) = 0 \forall i \neq j, i,\ j\ \in C)$.
A cell is occupied by at most one car.

ii) Let for $c \in C$:

$$B_c^* := \{(i,j) \in B | pos_K(c) \leqslant i \leqslant pos_K(c) + le_K(c), j = cr_K(c), \text{if } hv_K(c) = 1;$$

$$i = cr_K(c), pos_K(c) \leqslant j \leqslant pos_K(c) + le_K(c),\ \text{if } hv_K(c) = 0\}$$

. Then $\forall c \in C : o(c,b) = 1\ \forall b \in B_c^*$ und $o(c,b') = 0\ \forall\, b' \in\ B \backslash B^*$. A car occupies exactly 2 or 3 consecutive cells in a row or a column.

iii) there is a red car which needs to be removed: $r \in C$ with

$$\begin{aligned} hv_K(r) &= 1 \\ cr_K(r) &= 2 \\ pos_K(r) &= max\{pos_K(i), i \in C, hv_K(i) = 1 \text{ und } cr_K(i) = 2\} \end{aligned}$$

The car which needs to be removed, is horizontally placed in the row with index 2, and there is no further horizontally placed car right of it in the same row.

In the following, when a configuration is mentionned, a valid configuration is implicitly meant.

MOVE A transformation from a configuration v into a configuration w is called a *move*, if the following holds:

i) $\exists j \in C : pos_v(j) \neq pos_w(j)$

ii) $pos_v(i) = pos_w(i)\ \forall i \in C,\ i \neq j$

iii) $hv_v(i) = hv_w(i)\ \forall i \in C$

iv) $cr_v(i) = cr_w(i)\ \forall i \in C$

v) $le_v(i) = le_w(i)\ \forall i \in C$

vi) all configuration z are valid configurations with

$$z \in \{pos_v(j) < pos_z(j) < pos_w(j),\ j \in C,\ pos_v(i) = pos_z(i) \forall i \in C,\ i \neq j,$$

$$\forall i \in C : hv_v(i) = hv_z(i),\ cr_v(i) = cr_z(i),\ le_v(i) = le_z(i)\}$$

BLOCKING CARS We define the set $B_v^+(c)$ for a configuration v and a car $c \in C$ as the set of cars which block c from moving in positive direction (down or right):

$$\begin{aligned} B_v^+(c) = \quad \{c' \in C \mid \quad & (hv_v(c) = hv(c') \wedge \\ & cr(c) = cr(c') \wedge \\ & pos(c') > pos(c)) \vee \\ & (hv_v(c) \neq hv_v(c') \wedge \\ & pos_v(c') \leqslant cr_v(c) \leqslant pos_v(c') + le_v(c') \wedge \\ & pos_v(c) < cr_v(c'))\} \end{aligned}$$

Therefore, it is the set of cars which prohibit that car c can move in positive direction to the border of the board. $B_v^-(c)$ is defined similarly as the set of cars which block c from moving in negative direction. Then $B_v(c) = B_v^+(c) \cup B_v^-(c)$.

SOLUTION CONFIGURATION A configuration is called *solution configuration* or *final configuration*, if $B_w^+(r) = \emptyset$.

SOLUTION PATH A *solution path* is a sequence of states $v_0 v_1 \dots v_l$, whereas $v_i v_{i+1}$ is a legal move, v_l is a final configuration and v_i is not a final configuration $\forall\ i \in \{0, \dots, l-1\}$.

SET OF START CONFIGURATIONS The *set of start configurations* $\mathcal{S}$ contains all configurations for which there exists a solution path.

BASICS FROM GRAPH THEORY

UNDIRECTED GRAPH An *undirected graph* is a tuple $G = (V, E)$ with a set V (so-called nodes) and a set $E \subseteq V \times V$ (so-called edges) whereas E is a set of unordered pairs of nodes. If $e = (v, w) = (w, v) \in E$ with $v, w \in V$, we say that v and w are incident with e and call v and w neighbors or adjacent to each other.

DEGREE Let $v \in V$, then the *degree* of a node $deg(v)$ is defined as the number of its neighbors.

DIRECTED GRAPH A *directed graph* $G = (V, E)$ with node set V has an edge set $E \subseteq V \times V$ of ordered pairs of nodes, such that $(v, w) \neq (w, v)$ für $v, w \in V$. For an edge $e = (v, w)$ we call v the predecessor of w and w the successor of v. For a node v we distinguish between its *in-degree* and its *out-degree*: $deg_{out}(v)$ denotes the number of nodes for which v is predecessor; $deg_{in}(v)$ denotes the number of nodes for which v is successor.

PATH We call the sequence of nodes $v_0 \dots v_k$ with $v_i \in V, i = 0, \dots k$ and $e_i = (v_i, v_{i+1}) \in E$ with $i \in \{0, \dots, k-1\}$ a *path* with starting node v_0 and end node v_k. The path has the length of k.

PROBLEM SPACE The *problem space* or *state space* for a start configuration $s \in \mathcal{S}$ is denotes as a directed graph $G_s = (V, E)$, with V the set of all by legal moves from s reachable configurations, and E set of all legal moves. V can be seen as union of the disjoint sets $\{s\}$, F and I: F contains all final configuration which can be reached from s by legal moves, I contains all intermediate states. Therefore, a solution path is a path through the problem space with s as starting node and a configuration $f \in F$ as end node. We define the set $FF \subseteq F$ as set of final configurations whose distance to s is minimal (i. e.the length of the shortest path from s to $f \in FF$). All paths with starting node s, end node $f \in FF$ and of this length are called *optimal*. The out-degree of a node is the number of possible moves from this node.

GAME Given a starting configuration $s \in \mathcal{S}$. A *game*is the task to find a path in the corresponding problem space with starting node s and end node $f \in F$.

OPTIMAL PROBLEM SPACE The subgraph $G_o := (V_o, E_o)$ of G is called the *optimal problem space* if V_o contains all nodes and E_o contains all edges which are contained in an optimal path in the problem space.

COMPLEXITY MEASURE We define a complexity measure as a mapping $\mathcal{C} : \mathcal{S} \to \mathbb{R}$, which assigns each start configuration a real number.

COMPLEXITY MEASURES For a given start configuration $s \in \mathcal{S}$ and its corresponding problem space $G_s = (V, E)$, we define the following complexity measures:

SIZE OF STATE SPACE

- $\mathtt{nodes} = |V|$
- $\mathtt{nodf} = |V \setminus F|$
- $\mathtt{edges} = |E|$, whereas the reciprocal edges $e = (v, w)$ and $e' = (w, v)$ between two nodes $v, w \in V$ are counted once.

LENGTH OF SOLUTION PATH

- $\mathtt{lsp} = \min\{k|(s = v_0 \dots v_k)$ is optimal solution path$\}$
- $\mathtt{avlsp} = \frac{1}{|F|} \sum_{f \in F} \mathtt{lsp}(f)$,
  whereas $\mathtt{lsp}(f) := \min\{k|(s = v_0 \dots v_k = f)$ is optimal solution path$\}$ for any $f \in F$.

NUMBER OF DECISION POSSIBILITIES

- $\mathtt{avdg} = \frac{1}{|V \setminus F|} \sum_{v \in V \setminus F} \mathtt{deg_{out}}(v)$
- $\mathtt{avdgop} = \frac{1}{|V_o \setminus F_o|} \sum_{v \in V_o \setminus F_o} \mathtt{deg_{out}}(v)$
- $\mathtt{avdgog} = \frac{1}{|V_o \setminus F_o|} \sum_{v \in V_o \setminus F_o} |\{(v, w) \in E\}|$
- $\mathtt{br} = \frac{1}{|V_o \setminus F_o|} \sum_{v \in V_o} \mathtt{br}(v)$, whereas $\mathtt{br}(v) = 1 - \frac{|\{(v,w) \in E_o \mid w \in V_o\}|}{|\{(v,u) \in E \mid u \in V\}|}$ for a node $v \in V$.

NUMBER OF SOLUTION PATHS

- $\mathtt{sp} = |\{(v_0 \dots v_k)|v_0 = s, v_k \in F, (v_0 \dots v_k)$ is optimal solution path$\}|$
- $\mathtt{sppf} = \frac{\mathtt{sp}}{|F|}$

GAME PROPERTIES

- $\mathtt{cars} = |C|$
- $\mathtt{fields} = \sum_{c \in C} \mathtt{le}_s(c)$
- $\mathtt{mc} = \frac{1}{|V \setminus F|} \sum_{v \in V \setminus F} \mathtt{mc}(v)$, whereas $\mathtt{mc}(v)$ is the number of cars which can be moved at least one cell up/down/left/right in configuration v
- $\mathtt{mcop} = \frac{1}{|V_o \setminus F_o|} \sum_{v \in V_o \setminus F_o} \mathtt{mc}(v)$

COUNTERINTUITIVE MOVES A move from configuration v to configuration w is called counterintuitive, if $\mathtt{pos}_v(r) > \mathtt{pos}_w(r)$ or $|B_v^+(r)| < |B_w^+(r)|$, where $r \in C$ is the red car which needs to be removed.

- $\mathtt{cm} = \frac{\sum_{e \in E_c} \omega(e)}{\mathtt{sp}}$ with $\omega : E \to \mathbb{N}$, $\omega(e)$ indicates in how many optimal paths the edge e occurs, and $E_c = \{e \in E_o | e$ ist unintuitiv$\}$
- $\mathtt{cmpl} = \frac{\mathtt{cm}}{\mathtt{lsp}}$

	CORR		B	I	A	E	G
nodes	0.151	Mean	3613.4	3800.2	5562.0	6212.9	6380.0
		SD	4645.7	4364.6	5932.0	7779.5	6292.9
nodf	0.152	Mean	3430.0	3646.1	5367.6	5983.0	6125
		SD	4481.2	4257.3	5735.6	7508.8	5956.5
edges	0.110	Mean	19419.2	18332.6	27910.3	29349.7	29748.9
		SD	26817.8	25978.1	32968.3	41649.2	35887.9
lsp	0.794	Mean	10.1	21.50	25.77	31	37.25
		SD	3.71	7.59	6.68	7.3	3.18
avlsp	0.766	Mean	13.03	24.92	29.2	34.28	41.06
		SD	4.87	8.20	7.15	8.30	3.89
avdg	−0.108	Mean	9.32	8.35	8.88	8.27	8.44
		SD	2.39	1.99	1.96	1.85	1.54
avdgop	0.088	Mean	1.600	1.720	1.677	1.750	1.706
		SD	0.356	0.351	0.332	0.337	0.297
avdgog	−0.127	Mean	9.45	8.41	8.55	8.13	8.46
		SD	2.592	1.889	1.974	1.763	1.419
br	−0.136	Mean	0.808	0.779	0.781	0.766	0.786
		SD	0.058	0.054	0.053	0.053	0.032
sp	0.0701	Mean	989.1	2 708 525	30 051 880	1 055 018 000	40 683 410
		SD	2493.6	11 028 490	159 122 200	4 878 146 000	187 725 400
sppf	0.065	Mean	174.0	1 350 731	23 772 900	5.8	40 272 220
		SD	309.3	7 978 521	151 017 900	3 113 466 000	187 814 300
cars	0.399	Mean	8.6	11.04	11.31	11.98	11.88
		SD	2.59	2.37	1.81	1.79	1.08
fields	0.404	Mean	19.9	25.2	25.7	27.0	26.63
		SD	5.25	4.70	3.48	3.37	1.79
mc	0.149	Mean	4.89	5.15	5.42	5.46	5.37
		SD	1.41	1.11	0.99	1.22	0.72
mcop	0.125	Mean	4.773	5.03	5.12	5.29	5.29
		SD	1.48	1.15	1.01	1.11	0.75
cm	0.103	Mean	0.641	0.751	0.869	0.949	0.976
		SD	0.608	0.819	1.020	1.174	0.791
cmpl	−0.255	Mean	0.065	0.039	0.035	0.033	0.027
		SD	0.079	0.048	0.040	0.040	0.022

Table 3: Mean and standard deviation of the introduced measures, grouped by the games' difficulty rating (beginner (B), intermediate (I), advanced (A), expert (E), grand master (G)), as well as the measures' correlation to the game's difficulty (Pearson correlation coefficient).

A.2 *Values of the complexity measures*

A.3 *For the experiment selected games*

	NODES	EDGES	LSP	AVDG	BR	SP	SPPF	CARS	FIELDS	MC	MCOP	CM	CMPL
Mean	5245	25434	26.4	8.6	0.78	$20.77 \cdot 10^7$	$17.98 \cdot 10^7$	11.26	25.5	5.32	5.1	1.08	0.04
SD	5882	32628	9.96	1.95	0.05	$210.7 \cdot 10^7$	$134 \cdot 10^7$	2.14	4.2	1.09	1.08	1.05	0.05
B													
Mean	3613	19419	10.1	9.32	0.8	989.1	174.0	8.6	19.9	4.89	4.77	0.64	0.06
SD	4646	26817	3.7	2.4	0.06	2494	309	2.6	5.25	1.41	1.48	0.69	0.08
del 07B	7273	41384	12	12.1	0.88	576	18	9	18	6.6	6.7	1.1667	0.0972
1 03B	4821	25166	15	10.5	0.85	1536	384	10	21	6.43	6.06	0.9167	0.0611
del 01B	1075	5821	7	10.9	0.86	61	15.3	8	20	5.4	5.8	0	0
del 04B	451	2008	8	9.07	0.8	2	2	7	18	3.8	3.3	1	0.125
del 05B	2784	14786	8	10.66	0.8	575	288	11	26	6.0	6.2	0	0
del 02B	21055	119889	7	11.5	0.8	42	42	11	25	6.5	6.85	0	0
del 06B	2954	14047	8	9.58	0.87	6	6	11	26	5.48	6.25	1	0.13
1 10B	51	94	14	3.78	0.61	8	8	11	24	2.9	2.8	1.25	0.09
I													
Mean	4139	1806	21.5	8.3	0.78	$0.271 \cdot 10^7$	$0.135 \cdot 10^7$	11.0	25.22	5.2	5.0	0.75	0.04
SD	4364	25978	7.59	1.99	0.053	$1.103 \cdot 10^7$	$0.798 \cdot 10^7$	2.37	4.70	1.11	1.15	0.82	0.048
2 07I	3182	13013	24	8.83	0.78	$0.019 \cdot 10^7$	63840	12	28	5.65	5.79	0.3357	0.0140
3 08I	1338	4786	26	7.26	0.73	48840	16280	14	31	5.82	6.12	0.0020	0.0001
1 15I	1128	3751	14	6.70	0.77	228	228	14	32	5.67	4.85	1.5000	0.1071
1 19I	561	1604	40	5.90	0.79	123	24.6	11	25	3.7	3.96	5.5772	0.1394
A													
Mean	5562	27910	25.8	8.9	0.78	$3.01 \cdot 10^7$	$2.38 \cdot 10^7$	11.3	25.7	5.4	5.1	0.87	0.03
SD	5932	32968	6.68	1.96	0.053	$15.9 \cdot 10^7$	$15.1 \cdot 10^7$	1.81	3.48	0.99	1.01	1.0	0.04
1 26A	196	483	22	5.0	0.82	8	8	9	20	2.64	2.88	4.5000	0.2045
1 23A	4671	19068	29	8.58	0.74	3782	3782	14	30	6.12	5.08	1.4804	0.0510
4 18A	8052	40360	29	10.2	0.79	$0.21 \cdot 10^7$	21480	10	24	5.6	5.4	0.6562	0.0226
1 22A	530	1514	33	5.9	0.73	468	78	14	31	4.73	4.65	5.6667	0.1717
E													
Mean	6213	29350	31	8.27	0.77	$105.5 \cdot 10^7$	5.84	11.98	26.96	5.46	5.29	0.95	0.033
SD	7780	41649	7.26	1.85	0.05	$487.8 \cdot 10^7$	$311.3 \cdot 10^7$	1.79	3.37	1.22	1.11	1.17	0.04
del 39E	754	2231	32	6.0	0.71	21600	7200	12	27	4.39	4.10	0.0125	0.0004
2 28E	75	106	31	2.85	0.63	1	1	13	29	2.26	2.52	4.0000	0.1290
3 25E	6262	30793	29	10.41	0.82	18240	3040	11	27	5.98	5.58	0.9000	0.0310
1 32E	23009	113755	48	10.02	0.80	$2058 \cdot 10^7$	$2058 \cdot 10^7$	13	29	6.94	5.97	2.3959	0.0499
1 38E	3493	13866	50	8.02	0.70	$12.7 \cdot 10^7$	$2.1 \cdot 10^7$	13	29	5.74	6.66	0.2652	0.0053
G													
Mean	6380	29749	37.3	8.44	0.79	$4.1 \cdot 10^7$	$4.0 \cdot 10^7$	11.88	26.63	5.37	5.28	0.98	0.03
SD	6293	35888	3.18	1.54	0.03	$18.8 \cdot 10^7$	$18.8 \cdot 10^7$	1.08	1.79	0.72	0.75	0.79	0.02
2 37G	5824	23352	41	8.21	0.78	$0.11 \cdot 10^7$	276080	11	25	5.05	5.20	0.8796	0.0215
del 55G	1583	5129	41	6.78	0.77	2688	448	11	25	4.32	3.95	2.5268	0.0616
2 36G	1062	2992	38	5.76	0.75	12320	3080	12	26	3.93	4.03	0.1312	0.0035

Table 4: The measure values of the 24 games which were selected for the experiment. Striking values are highlighted in red. Mean value and standard deviation (SD) are shown for each complexity class.

[1] David Anderson, Emily Anderson, Neal Lesh, Joe Marks, Brian Mirtich, David Ratajczak, and Kathy Ryall. Human-guided simple search. In *AAAI/IAAI*, pages 209–216, 2000.

[2] J. R. Anderson. Problem solving and learning. *American Psychologist*, 48, 1993.

[3] John R. Anderson. *Cognitive Psychology and its Implications*. W. H. Freeman and Co., New York, NY, 1980.

[4] G. S. Halford, W. H. Wilson, and S. Phillips. Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behavioral & Brain Sciences*, 21:803–865, 1998.

[5] Kenneth Kotovsky, John R Hayes, and Herbert A Simon. Why are some problems hard? evidence from tower of hanoi. *Cognitive psychology*, 17(2):248–294, 1985.

[6] M. Ragni, F. Steffenhagen, and T. Fangmeier. A structural complexity measure for predicting human planning performance. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 2012.