Analysis of human problem solving drafts: a methodological approach on the example of Rush Hour

Mareike Bockholt[1], **Olaf Peters**[2], **Susanne Narciss**[2], **Katharina A. Zweig**[1]
[1] Graph Theory and Complex Network Analysis Group, Department of Computer Science,
University of Kaiserslautern, Germany
[2] Chair of Psychology of Learning and Instruction, Department of Psychology,
Technical University of Dresden, Germany
{mareike.bockholt, zweig}@cs.uni-kl.de, {olaf.peters, susanne.narciss}@tu-dresden.de

Abstract

Assessing the quality of a learner's solution for a given task is an essential step in analyzing a learner's performance. For a well-defined sequential problem, correctness and optimality of the solution as well as its length provide first simple and reasonable metrics. However, this ignores the fact that there are conceptually different errors that humans make when solving a problem. This work proposes a rule-based system of error categories which is able to classify conceptually different errors with respect to their (assumed) motive. The principles the categories are based on are valid for most well-defined sequential problems and can hence serve as a valuable tool in the analysis of human solutions for such a problem. In this work, the error category system is adapted to the game Rush Hour. We use the category system as a tool for a detailed analysis of 115 human solutions of a Rush Hour game. We found that the most common error type is based on a simple solving heuristic, but mainly occurs in the first half of the solution process. Other error types whose occurrence is numerically less dominant, are still found in the majority of the solutions. However, they occur in very specific game situations. As a first generalization approach of the category system, its application on a further dataset containing 56 different Rush Hour tasks and more than $31,000$ human solutions yield promising results.

Keywords: Problem solving; Solution quality; Error analysis; Error categories; Rush hour

Introduction

Assessing the quality and characteristics of the solution of a learner for a given task is an essential step in problem solving research (Newell & Simon, 1972; Anderson, 1993). Due to the complexity of the task, there are often no precise or only very simple quality measures. For well-defined problems, i.e., when a well-defined initial situation needs to be transformed into a well-defined final situation by applying operators from a fixed set of operators, solving success (was the final situation reached?) or solving time (how long did it take?) are simple and natural metrics. For well-defined sequential problems, i.e., problems that cannot be solved in a single step, but require a sequence of actions, optimality or length of the solution are reasonable metrics: a shorter solution is better than a longer one. This can even be considered at the level of single moves: every move is either correct, wrong, or unnecessary. However, this ignores the fact that there are conceptually different kinds of errors that can be made by the problem solver: For instance, an error may be due to a false idea of the final configuration, to incomplete knowledge of the available operators, or to an accidentally wrong action. There is hence a need for a categorization of possible errors that allows distinguishing what *kind* of error was made by the problem solver. This can help to understand underlying problem-solving processes, to identify situations within the solution process that are prone to certain types of errors, or generally allows a more detailed analysis of human solutions for a task. At a task-independent level, Norman and Reason, distinguish human errors into *planning failures* (mistakes) and *execution failures* (slips) (Reason, 1990): "An error in the intention is called a mistake. An error in carrying out the intention is called a slip." (Norman, 1983) While this is a helpful general categorization, an error type categorization obviously needs to be more detailed in order to understand the problem-solving process.

Analysis and domain-specific categorization of human errors was done in various contexts several decades ago: types of spelling errors (Caramazza, Miceli, Villa, & Romani, 1987), types of human errors in man-machine interaction (Rasmussen, 1982), errors in patient medication in hospitals (Leape, 1994; Zhang, Patel, Johnson, & Shortliffe, 2004), conceptual errors when learning mathematical concepts (Eichelmann, Narciss, Schnaubert, & Melis, 2012), or performing a proof (Autexier, Dietrich, & Schiller, 2012). For a well-defined sequential problem, however, very few approaches exist for categorizing possible errors, such as error moves in the block design puzzle (Toraldo & Shallice, 2004).

To accomplish this aim, this work provides a set of eleven error types, referred to as a *category system*, that aims at covering the majority of errors made by humans when solving a task. The categories are based on principles valid for most types of well-defined problems. It is rule-based and allows an automatic classification of error moves. It can be used as an additional analysis tool for human solutions in different experimental settings and for various research questions.

In this work, the categories are applied and tested on one specific task: a sliding-block puzzle called Rush Hour (invented by Nob Yoshigahara, distributed by ThinkFun Inc. and by HCM Kinzel). Two example Rush Hour levels are shown in Figure 1. Blocks, representing cars, are placed on a board with one designated exit on the right, representing a parking lot. Cars can only be moved forwards or backwards, but not sideways. The goal is to remove the target car (dark color) from the board through the exit.

Rush Hour is a well-defined sequential problem and can be understood as a transformation problem, according to the problem classification by Greeno (1978). Although all el-

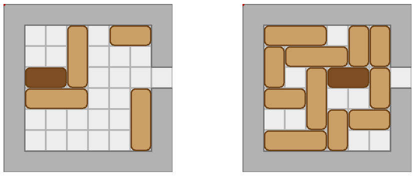

Figure 1: Two example levels of Rush Hour. Left: low difficulty (solvable in 10 moves); right: medium difficulty (solvable in 25 moves). The latter was used as experimental level in this study.

ements, rules, and principles are transparent to the learner, solving a transformation problem can be difficult for humans; this is also true for Rush Hour (Ragni et al., 2011; Bockholt & Zweig, 2015).

For Rush Hour, too, different wrong or unnecessary moves can be of a different "nature": Bennati, Brussow, Ragni, and Konieczny (2014) show that the presence of so-called clusters in a Rush Hour configuration, i.e., two cars placed directly next to each other, decreases the chance for a player to find an optimal solution. They explain this with Gestalt effects: humans tend to perceive the two cars as one object and tend to move them together—although this might be wrong or unnecessary. This is an example of one "type" of errors.

In this paper, we apply the derived category system on the solutions of 115 participants solving one single Rush Hour task and answer the following questions: (i) Which error types are most common in a Rush Hour solution? (ii) Do the different error types occur in arbitrary phases of the solution process? In order to generalize the category system, we used a different dataset containing 56 different Rush Hour tasks and more than $31,000$ human solutions for those tasks in total (collected by Jarušek and Pelánek (2012)).

Error category system

Every configuration of a well-defined problem induces a finite *problem space*, containing all board configurations reachable from the initial one by allowed moves. If a problem space contains a configuration in which the goal is reached (also called *final state*), the configuration is solvable. In a Rush Hour game, all moves are reversible, hence, a final state can be reached from every state in the problem space (if the initial configuration is solvable). This is different for other tasks where one wrong move can lead the player to a "dead end" from which no final state can be reached anymore. This implies that for the problem space of any (solvable) Rush Hour configuration: (i) a move can increase or decrease the distance to the closest final state by at most one step; (ii) in every (non-final) state, there exists at least one *correct* move that decreases the distance to the closest final state by one step. Besides correct moves, there might exist *wrong* moves that increase the distance to the closest final state by one step, and *unnecessary* moves that neither increase nor decrease this distance. Both–wrong and unnecessary moves–will be called *error moves* in the following since a solution containing at least one of them cannot be optimal anymore. Note that an unnecessary move increases the length of the solution by exactly one step while a wrong move increases it by two steps because it needs to be corrected by an additional move.

However, error moves may occur for different reasons. We hypothesize different types of errors exist that are conceptually different. The following paragraphs will introduce eleven different error types: four have a descriptive quality, while seven are based on the actual game situation (*conceptual error categories*) and assume a kind of motive that leads to this type of error. For any error move found in a solution, whether wrong or unnecessary, we can then use the categories' rules to automatically check in which category it falls.

Descriptive error categories The following four categories describe which pattern the error shows. They are applicable to any solution of any sequential problem. A move from one board configuration v to another configuration w will be denoted as $v \to w$. If a player's solution contains the moves $v \to w$ as well as $w \to v$, at least one of them is an error move. In such a case, the error move is categorized as a ***Generalized Undo Mistake***. If the moves $v \to w$ and $w \to v$ are directly consecutive in a player's solution, the error move is additionally categorized as an ***Undo Mistake***. Furthermore, it is never optimal to perform two consecutive moves which effect could have been reached by only one move. In the case of Rush Hour, this includes consecutive moves of the same car into the same direction (up/down/left/right). Such error moves are called ***Two In One Mistakes***.

There are error moves that are unnecessary because the move "does not make a difference". Consider, for example, the configuration on the left in Figure 1: Moving the horizontal car of length 3 to any other position does not change the game situation because in all its possible positions, the car blocks the same set of other cars in their movement. This is the idea of equivalent game situations. An error move $v \to w$ is classified as an ***Equivalence Trap*** if v and w are equivalent. While the definition of equivalent game situations is task-dependent, the error type Equivalence Trap is generally applicable on any sequential problem. In the case of Rush Hour, for a given board configuration, we say that a car i blocks a car j if i occupies at least one board cell in the same row as j (if car j is horizontal) or in the same column as j (if j vertical). For the board configuration shown in Fig. 1 on the right, the target car is blocked by three vertical cars. Two Rush Hour board configurations are then called equivalent if each car blocks exactly the same set of cars in both configurations.

Conceptual error categories The descriptive error categories can be seen as symptoms of an erroneous solving approach: the presence of such error types is a sign for an erroneous solution approach, but they don't allow any insight into *why* the error move occurred. The following *conceptual error categories*, on the other hand, are rooted in a kind of conceptual error.

For most sequential tasks, there are simple solving heuristics that might be used by humans in the solving process.

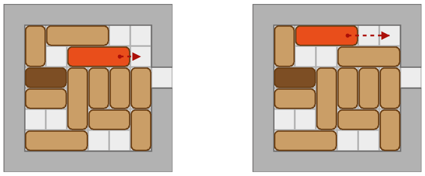

Figure 2: A Relaxed Car Unit Mistake: the left move is correct, followed by a wrong move (shown on the right) that is classified as a Relaxed Car Unit Mistake.

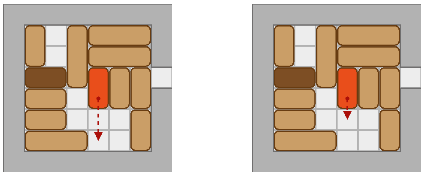

Figure 3: The (wrong) move shown on the left is classified as a Border Attraction, but not as an Early Unblock Mistake. The correct alternative move is shown on the right.

However, following a heuristic might provoke an error because the correct move conflicts with the heuristic.

Generally, any well-defined problem has a well-defined goal to be reached. Applying the hill-climbing heuristic (Newell & Simon, 1972), moves that seem to make the current situation more similar to the goal situation will be preferred, and moves that seem to make it less similar will be avoided. However, this might be a source of errors: seemingly correct moves might be performed too early in the course of the game, and counterintuitive moves necessary for reaching the goal are avoided. For Rush Hour, in order to reach the goal, the cells between target car and exit need to be freed, and the target car needs to be moved through the exit. Both versions of phrasing the goal of the game might provoke "Avoid"- and "Too-Early"-mistakes.

A move which is actually blocking the exit is seemingly a detour and might be avoided. An error move is called an ***Avoid Blocking The Exit Mistake***, if (i) neither before nor after the move, the moved car is on any cell to the right of the target car, and (ii) all correct alternative moves require moving the same or any other car on a cell to the right of the target car. In other words, in order to come closer to the final state, it is necessary to block the exit, but the player avoids this and performs a wrong or unnecessary move instead.

As a corresponding "Too-Early"-mistake, we call an error move an ***Early Unblock Mistake*** if (i) the moved car blocked the exit before the move, but (ii) does not block it after the move, and (iii) no correct alternative move exists where the same car is moved in the same direction. The last requirement ensures that an error move as shown in Fig. 3 (left) is not classified as an Early Unblock Mistake since there is a correct alternative move that also frees the exit (shown on the right).

Similarly to the premature unblocking of the exit, the same behavior is captured for the premature movement of the target car towards the exit. An error move is categorized as a ***Early Target Car Move*** if the target car is moved towards the exit (despite not being beneficial in this game situation).

Similar to Avoid Blocking The Exit mistakes, moving the target car to the left, i.e., further away from the exit, might seem to be a detour and might be avoided. Hence, an error move is categorized as ***Avoid Moving Target Car Backwards*** if (i) it does not involve the target car being moved to the left, but (ii) all correct alternatives do. In other words, it would have been necessary to move the target car further away from the exit, but this is avoided by the player.

While the previous four categories are based on conflicts with the hill-climbing heuristic, the following three might be explained by conflicts with heuristics for complexity reduction. One such heuristic, applicable on tasks with a spatial representation, is to consider structures locally and resolve them one after the other. This is sometimes a good strategy, but it often fails. Hence, an error might occur because the same local structure as in the previous move is considered, but a change to a different local structure would have been necessary. For Rush Hour, this corresponds to the simple procedure: check which cells have been freed by the previous move, and then move a car onto a previously freed cell. For Rush Hour, an error move $m = v \rightarrow w$ moving car i is a ***Stay Local Mistake*** if (i) in situation w, car i occupies at least one board cell that was freed by the move directly before m, and (ii) in situation v, there exists no correct alternative move that also uses one of those freed cells.

Treating two or more objects as a single one can be a further (possibly unintentionally applied) heuristic for complexity reduction. Bennati et al. (2014) report that the existence of so-called clusters of cars in a Rush Hour configuration has a significant effect on the optimality of solutions. Two cars, both horizontal or both vertical, are called a *cluster* or *car unit* if they are placed in two consecutive rows or columns, and if at least one cell occupied by the first car is adjacent to at least one cell occupied by the other car. For example, in Fig. 1 on the right, the two cars of length 3 in the first and second row are a cluster, while the two horizontal cars in the fifth and sixth row are not. Such a cluster might provoke error moves since the two cars might be perceived as a unit and be moved in consecutive moves, although only one of the cars needed to be moved. The same principle—different objects are perceived and treated as a unit—occur also in other well-defined problems. For Rush Hour, we call an error move $m = v \rightarrow w$ a ***Relaxed Car Unit Mistake***[1] if there is another move m' directly before or directly after m, hence $m' = v' \rightarrow v$ or $m' = w \rightarrow w'$, such that (i) in the configurations before and after m and m', there exists a cluster of cars; (ii) m and m' move those cars in the same direction; (iii) in the configuration before the moves m and m', there exists no alternative sequence of two correct moves involving the same two cars. Figure 2 shows an example: The move shown on the left is correct, the move on the right is wrong. In order to reach the next sub-goal of the game (moving the horizontal car in

---

[1] The prefix *relaxed* illustrates that a "perfect" cluster where the cars are perfectly aligned is not required.

the fifth row onto the free cells at the left border), the move shown on the left would have been sufficient.

Specifically for Rush Hour, it is cognitively less complex to move a car as far as possible than to evaluate how many cells are actually necessary. This strategy might provoke errors: An error move is categorized as a ***Border Attraction Mistake*** if (i) the moved car is moved as far as possible (constrained by the border of the board or by another car), and (ii) there exists a correct alternative that moves the same car in the same direction, but not as far as the error move. This means that the decision of moving the particular car is correct, but moving it that far is erroneous. An example is shown in Fig. 3: the left move is an error move and categorized as a Border Attraction Mistake because it would have been correct to move the same car by only one cell (shown on the right).

Experiment

A study with 138 participants was conducted at a German university (data of 23 participants had to be excluded due to incomplete data). Hence, the experimental data of 115 participants (95 female, 18 male; age: 18 to 30 years, $M_{age} = 20.88 \pm 2.80$) are used in the following analysis. All participants attempted to solve two rather easy exercise levels of Rush Hour in order to get acquainted with the rules of the game. After the exercise phase, all participants attempted to solve a Rush Hour game of medium difficulty, shown in Fig. 1 on the right (length of optimal solution: 25 moves), referred to as the experimental level.

The experimental level had to be solved within 10 minutes and 60 moves. The participants were allowed to undo their last move (but only exactly one, until another regular move is performed. Reversing the last move by a regular move is obviously always possible). Moves undone by a participant via the Undo-button were not included in the participant's solution in order to prevent the inclusion of accidental moves.

Results

We found that only 70 participants managed to solve the experimental level (45 failed). Not a single participant was able to solve the level within the optimal number of steps: the shortest solution found (by exactly one participant) contained 27 moves (optimal: 25). This shows that the proportion of participants able to solve the game at all or optimally, is not a sufficient metric for a detailed analysis. Overall, the participants' solutions of the experimental level contained an average of 18 ± 11 error moves, i.e., on average 8 ± 4 unnecessary moves plus 10 ± 8 wrong moves.

Error types in the solutions The proposed error category system was applied to each participant's solution of the experimental level: for each error move, we checked in which categories it falls. Out of the 2,134 error moves in total, 7 % do not fall in any category, 35 % into exactly one, and 36 % (18 %, 4 %, < 1 %) into 2 (3,4,5) categories. Although no distinction is made between conceptual and descriptive categories, the proportion of uncategorized error moves and the proportion of error moves belonging to multiple categories is reasonably small. This shows that the category system is a valid partitioning of typical error moves in this task. Figure 4 shows for each error category how often it occurs in the participants' solutions. The error types Stay Local and Generalized Undo are the most common errors. However, Relaxed Car Unit, Early Unblock and Avoid Blocking The Exit also occurred about three times on average in every solution. The error type Equivalence Trap did not occur at all in the solutions of this game (it does, however, occur in the solutions of the other dataset used for generalization, see below).

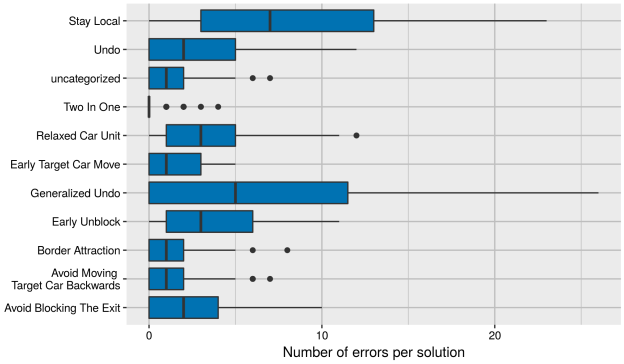

Figure 4: For each error category, the number of error moves of this type found in the participants' initial solutions is shown.

Table 1: For each error type, the number of the participants who made at least one such error in their solution is shown.

Stay Local	110/115	Border Attraction	66/115
Relaxed Car Unit	106/115	Undo	83/115
Early Target Car Move	76/115	Early Unblock	94/115
Generalized Undo	84/115	Two In One	23/115
Avoid Moving Targ. Car Backwards	59/115	Avoid Blocking The Exit	77/115

It is worth considering how many of the participants managed to solve the game without any occurrence of errors of a particular category. Table 1 shows the number of the solutions in which a certain error type is present a least once. Only 5 out of 115 participants solved the experimental level without any Stay Local mistakes. Also, almost all participants failed in solving the game without any Relaxed Car Unit Mistake. Furthermore, although the presence or error moves of the type Border Attraction, Avoid Moving Target Car Backwards, or Early Target Car Move does not appear dominant with respect to their total number of occurrence (cf. Fig. 4), at least half of the participants had at least one such error in their solution.

Location of errors in the solutions In addition to the frequency of occurrence of the error types, we considered the location of each error type in the participants' solutions. We computed the number of occurrences of each error type dependent on the location in the solution; more precisely, dependent on the number of steps necessary to reach the final state from the configuration in which the error move was made. This has the advantage that the location of errors in an almost optimal solution can be compared to that in a long

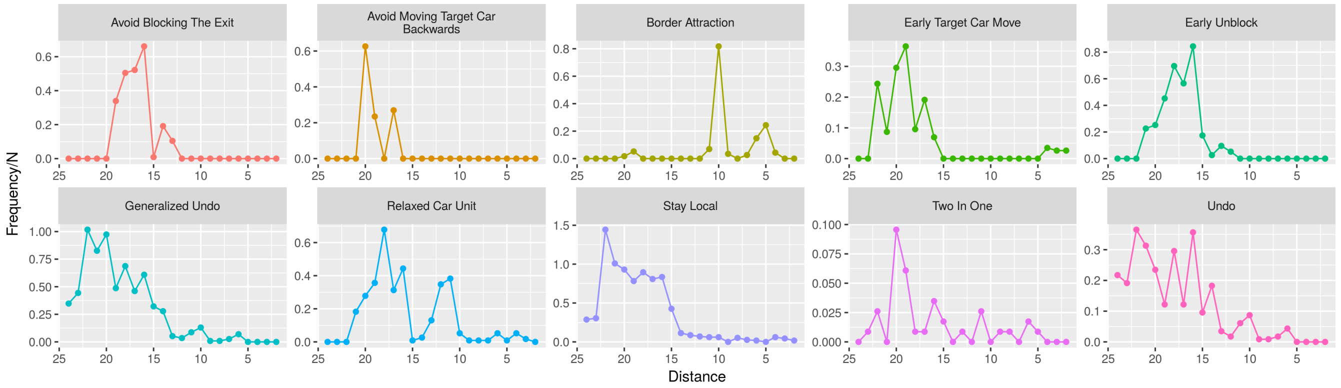

Figure 5: Counting the number of error moves of each type occurring at each distance from the goal. *x*-axis: remaining steps necessary for reaching the goal (25 in start configuration to 0 in final configuration). *y*-axis: total number of error moves of the corresponding type that occurred at the corresponding distance, normalized by the number of participants.

solution containing a large number of moves that do not decrease, or even increase, the distance to the final state. For the experimental level, the distance to the goal from the start configuration is exactly 25 steps. The problem space does not contain any state with a distance to the goal larger than 25 steps. Figure 5 shows the number of error moves of each type for each distance to the final state, normalized by the number of participants.

It seems that there are two patterns of error occurrence: Some occurred in specific game situations (or at least at very specific distances) while others occurred independent of the game situation, but only in the first half of the solution. Most of the conceptual error types show the former behavior: especially for Avoid Blocking The Exit, Avoid Moving Target Car Backwards, Border Attraction, and Relaxed Car Unit, the peaks of occurrence are narrow. The error types Early Unblock, Generalized Undo, Early Target Car Move, Undo and Stay Local, do not appear to depend on the situation. Interestingly, they rather occurred in the first half of the solution and only rarely in the second half. There are three possible explanations for this effect: (i) At the beginning of the solution process, the participants follow an exploratory "trial-and-error" approach, which is replaced by a rather target-oriented approach in the further course of the game due to learning effects during the solving process. (ii) Planning the next moves becomes easier when the goal state is "within sight". As soon as the distance to the goal is small enough such that the appropriate steps needed to reach it can be planned by the human, the number of errors decreases. However, their number dropped in the distance range of 10 to 15 remaining steps, which—we assume—seems to be too many for planning ahead. (iii) The reason may be related to the actual situation on the board: Since we only considered the solutions for one specific Rush Hour game, it may be that certain error types are actually only *possible* in certain game situations and not in others.

Generalization of category system

The previous section showed that the proposed category system is a useful tool for analysing Rush Hour solutions—at least for the one task used in the experiment. As a first approach for investigating whether the category system is general enough for the application on other Rush Hour tasks, we used 56 different Rush Hour tasks (levels) and players' solutions for these as a test dataset. In a further step, the applicability on other tasks needs to tested. The data used here was collected by Jarušek and Pelánek (2012), who developed a web-based tool for learning by problem-solving that is used in educational contexts. Unsolved solution attempts and solutions that took longer than three times the optimal solution were excluded from the analysis. The remaining data comprises 56 tasks with an optimal solution length between 3 and 50 moves (median 29), with 13 to 2,319 solutions for each task (median 160); more than 31,000 solutions, in total.

In the 31,000 solutions, more than 190,000 error moves were observed; the mean number of error moves per solution highly depended on the task, ranging from 2.1 ± 1.2 error moves for a task with an optimal solution length of 3, to 41.3 ± 12.5 for a task with an optimal solution length of 47. Although this dataset includes games played without full attention or with no incentive for finding the shortest possible solution, it is suitable for testing whether the proposed error category system is a reasonable approach for typical errors in Rush Hour solutions.

Table 2 shows the proportions of the observed error moves categorized into the different categories. 23 % do not fall into any category, while 60 % can be explained only by the conceptual categories. Especially at the easier levels, there are moves that cannot be categorized: If only tasks with an optimal solution length of at least 13 are considered, the proportion of uncategorized error moves drops to 17 %.

Table 2 also shows that only 16 % of all error moves are categorized into more than one conceptual error category, and only 1 % fall into more than two categories. The multiple classifications are mainly due to the Stay Local mistake: If we exclude this category, the percentage of moves classified into more than one conceptual category drops to 6 %. Hence, the number of error moves classified into more than one conceptual category is reasonably small.

It can be summarized that although the category definitions

Table 2: Application of the error category system on a larger set of tasks: Results of the categorization of the 194,290 errors in the players' solutions of all tasks, and of the 89,540 errors in the solutions for tasks with an optimal solution length at least 13. It is shown how many of the error moves fall into how many error categories.

	in ... categories			in ... conceptual cat.			
Tasks	0	>1	>2	0	1	>1	>2
all	23 %	41 %	13 %	40 %	44 %	16 %	1 %
less easy	17 %	44 %	16 %	33 %	48 %	19 %	2 %

are rather strict, the variety of different game situations is large, and there are only seven conceptual error categories, the introduced error category system is well generalizable on the test game data. Still, about a quarter of all observed error moves cannot be explained by any category.

Conclusion

In this work, a category system for "typical" error moves in well-defined sequential problems is introduced. This is possible because for each move, it can be easily determined whether it is a correct move or not. The category system with eleven categories is rule-based, and allows the automatic classification of error moves with respect to their (assumed) motive. The approach of categorizing the errors allows a detailed analysis of human solutions for a task that is more informative than only considering the solving success, optimality, or length of the solution. It can thus be used as an analytical tool when human performance in solving a task is of interest.

In this paper, the error category system was adopted and implemented for the sliding-block puzzle Rush Hour. It is then used for the analysis of the solutions of 115 participants playing a Rush Hour game with a medium level of difficulty. We found that the most common error type in the participants' solutions was the Stay Local Mistake, which assumes a simple heuristic as a solving strategy: always use the previously freed space. Another common error was the Generalized Undo error, which implies that it happened quite frequently that the participants returned to a situation they had visited before. However, we found that these error types (Generalized Undo as well as Stay Local) mostly occurred in the first half of the solution process and very rarely in the second half.

Other error types showed a different pattern of occurrence: Error types such as Relaxed Car Unit, Border Attraction, Avoid Blocking The Exit or Early Unblock did not occur in large numbers in the solutions, but were present in the majority of the solutions. We found that these error types occurred only in very specific game situations.

A first generalization of the error category system on a dataset containing 56 Rush Hour tasks and more than 31,000 solutions showed that the category system is plausible. Still, a non-negligible number of errors was left uncategorized in the test dataset which needs to be explored further in future work.

The results presented in this paper are intended to demonstrate the potential of a detailed error analysis in human problem-solving drafts. For future work, it might be interesting to use it to compare several solutions (from one or several persons): Do all persons make the same errors or are there individual differences? Which error types are recognized as a mistake by the person and can therefore be avoided in a new attempt, and which error types will remain? A different aspect concerns the reason why errors occur: Are there game situations that are more prone to certain error types than others? Which characteristics can be identified that trigger the occurrence of an error?

References

Anderson, J. R. (1993). Problem Solving and Learning. *American Psychologist*, *48*.

Autexier, S., Dietrich, D., & Schiller, M. (2012). Towards an intelligent tutor for mathematical proofs. *arXiv preprint arXiv:1202.4828*.

Bennati, S., Brussow, S., Ragni, M., & Konieczny, L. (2014). Gestalt effects in planning: Rush-hour as an example. In *Proceedings of the Cognitive Science Society* (Vol. 36).

Bockholt, M., & Zweig, K. A. (2015). Why is this so hard? Insights from the state space of a simple board game. In *Joint International Conference on Serious Games* (pp. 147–157).

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition*, *26*(1), 59–85.

Eichelmann, A., Narciss, S., Schnaubert, L., & Melis, E. (2012). Typische Fehler bei der Addition und Subtraktion von Brüchen–Ein Review zu empirischen Fehleranalysen. *Journal für Mathematik-Didaktik*, *33*(1), 29–57.

Greeno, J. G. (1978). Natures of problem-solving abilities. *Handbook of learning and cognitive processes*, *5*, 239–270.

Jarušek, P., & Pelánek, R. (2012). Analysis of a Simple Model of Problem Solving Times. In Cerri, Stefano A. and Clancey, WilliamJ. and Papadourakis, Giorgos and Panourgia, Kitty (Ed.), *Intelligent Tutoring Systems* (Vol. 7315, p. 379-388). Berlin Heidelberg: Springer.

Leape, L. L. (1994). Error in medicine. *Journal of the American Medical Association*, *272*(23), 1851–1857.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Prentice Hall, Englewood Cliffs.

Norman, D. A. (1983). Design Rules Based on Analyses of Human Error. *Commun. ACM*, *26*(4), 254–258.

Ragni, M., Steffenhagen, F., & Fangmeier, T. (2011). A Structural Complexity Measure for Predicting Human Planning Performance. In *Proceedings of the Cognitive Science Society* (Vol. 33).

Rasmussen, J. (1982). Human errors. A taxonomy for describing human malfunction in industrial installations. *Journal of occupational accidents*, *4*(2-4), 311–333.

Reason, J. (1990). *Human error*. Cambridge university press.

Toraldo, A., & Shallice, T. (2004). Error analysis at the level of single moves in block design. *Cognitive neuropsychology*, *21*(6), 645–659.

Zhang, J., Patel, V. L., Johnson, T. R., & Shortliffe, E. H. (2004). A cognitive taxonomy of medical errors. *Journal of Biomedical Informatics*, *37*(3), 193–204.