EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Eye movements support mental processing in Tower of Hanoi problem solving

Masterarbeit
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Marcel Dorer
Tübingen, September 2016

Eidesstattliche Erklärung

Hiermit erkläre ich,

- dass ich diese Arbeit selbst verfasst habe.

- dass ich keine anderen als die angegebenen Quellen benutzt und dass ich alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe.

- dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist.

Tübingen, den 06.09.2016

First Reader:
Prof. Dr. Hanspeter A. Mallot
Cognitive Neuroscience
Department of Biology
Eberhard Karls Universität Tübingen
Germany

Second Reviewer:
Dr. Gregor Hardieß
Cognitive Neuroscience
Department of Biology
Eberhard Karls Universität Tübingen
Germany

Acknowledgment

First of all, I want to thank Prof. Dr. Hanspeter A. Mallot for accepting my request to work on this subject as well as for all the support and materials he provided. During my studies Prof. Dr. Mallot proved to be an outstanding, patient and understanding teacher whom I wish only the best for the future.

I want to specially thank Dr. Gregor Hardieß, who took the part of second reviewer and supervisor, for numerous productive discussions and most notably for his support during the experimental design, data analysis as well as programming phase of my work.

Furthermore, I want to thank the whole staff of the Cognitive Neuroscience lab for their continuous support, many productive discussions and a very comfortable working environment.

Last but not least I want to thank all the participants of this study, as well as of the preliminary experiment and the friends and colleagues, who sacrificed their free time for proofreading.

Abstract

Goal orientation is a common characteristic of human behavior, which is closely linked to problem solving in novel situations. Oftentimes, efficient problem solving performance indicates mental planning of appropriate behavior prior to the execution by mentally denoting a concept as well as the associated outcome (Goel [2002]). Regularities of such cognitive processing might be reflected in overtly observable eye movement patterns (Nitschke et al. [2012]), which can easily be observed by tracking eye movements during problem solving tasks. One type of such tasks are disk transfer tasks like the Tower of Hanoi, of which a modified version was used in this study. During the planning process of this task eye movements were recorded and compared to a very basic strategy. This strategy postulates that every action that is planned requires a fixation of the area in which the action takes place. Over the course of this experiment this strategy was found in more than one third of the trials, with an increased usage in trials of higher complexity. In addition, this study found that there is a lower number of errors in trials where this strategy was used. Overall, a high variance in a wide range of parameters was found across participants, which indicates the usage of multiple different eye movement strategies. Those different strategies need to be subject of additional research.

Kurzfassung

Zielorientierung ist ein gemeinsames Merkmal menschlichen Verhaltens, welches in engem Zusammenhang zum Problemlösen innerhalb neuer Situationen steht. Häufig deutet effizientes Problemlösen auf mentales Planen von angemessenem Verhalten, bei dem, vor der eigentlichen Durchführung, mental ein Konzept sowie das entsprechende Ergebnis aufgestellt wird, hin (Goel [2002]). Regelmäßigkeiten innerhalb solcher kognitiven Verarbeitungen könnten in offen beobachtbaren Augenbewegungsmustern (Nitschke et al. [2012]), welche mit Hilfe von Blickregistrierungssoftware einfach während Problemlösungsaufgaben aufgezeichnet werden können, wiedergespiegelt werden. Eine Art solcher Aufgaben sind Scheibenverschiebungsaufgaben wie die Türme von Hanoi, wovon eine modifizierte Version innerhalb dieser Studie verwendet wurde. Während des Planungsprozesses dieser Aufgabe wurden Augenbewegungen aufgezeichnet und mit einer sehr einfachen Strategie verglichen. Diese Strategie geht davon aus, dass jede geplante Aktion eine Fixierung des Gebietes voraussetzt, in dem die Aktion ausgeführt wird. Während dieser Studie konnte der Einsatz dieser Strategie in mehr als einem Drittel der Durchgänge bestätigt werden, hierbei erhöhte sich die Anwendung mit der Komplexität der Durchgänge. Zusätzlich zeigte diese Studie, dass die Fehlerzahl der Probanden in Durchgängen in denen die Strategie eingesetzt wurde geringer ausfiel. Insgesamt wurde zwischen den Probanden eine hohe Varianz in einem Großteil der Parameter festgestellt, was darauf hindeutet, dass mehrere verschiedene Augenbewegungsstrategien Anwendung fanden. Diese unterschiedlichen Strategien müssen noch Thema weiterer Untersuchungen sein.

Contents

1. Introduction

1.1. Problem solving and disk transfer problems

Although goal orientation is a common characteristic of human behavior, goals often cannot be achieved directly because previously approved behavioral schemata might not be suitable or applicable (Ward and Morris [2005]). Such novel situations lead to problems that can be resolved in different ways, the simplest being a problem solving behavior solely relying on trial and error. In many cases efficient performance serves as an indicator for an identification of appropriate behavioral sequences by means of mental planning prior to the execution. Therefore, mental planning is a form of problem solving, which denotes the mental conception and behavioral evaluation as well as the associated outcomes before their execution (Goel [2002]).

A large portion of past research on planning focused on disk transfer tasks like the Tower of London or Tower of Hanoi (figure 1.1.1.). These tasks consist of well defined, knowledge-lean problems with definite solutions. Within such tasks, experimental variation of the problem structure allows systematic manipulations of planning demands and task complexity (Kaller et al. [2004]). While such tasks are commonly used for neuropsychological assessments, for example for patients with dysfunctions and frontal lobe lesions (Cockburn [1995]), schizophrenia (Morris et al [1995]) or Parkinson disease (Dagher et al. [1999]), they also find application in cognitive research, especially for behavioral and neuroimaging studies (Morris et al. [1993], Baker et al. [1996] and Lazeron et al. [2000]). For these purposes, a broad range of different variations of disk transfer tasks were developed.

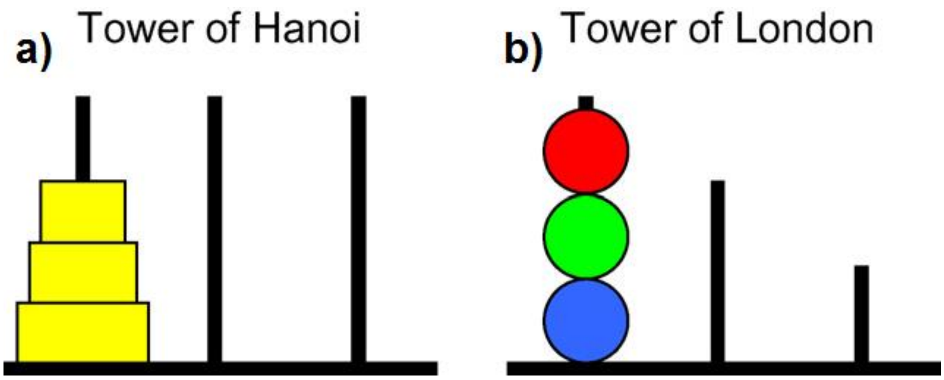


Figure 1.1.1.: Basic configurations of the Tower of Hanoi (a) and the Tower of London (b). Both tasks require to efficiently transform a given starting state into a specified goal state. The Tower of Hanoi (a) consists of three or more disks of different size that are distributed across three equally sized rods. The goal is to find the optimal solution with the minimum number of moves while moving one disk at a time and without placing larger disks on top of smaller ones. In contrast, the original version of the Tower of London (b) consists of three differently colored balls placed on three rods of different heights. A large variety of Tower of Hanoi and Tower of London tasks found application in past research.

Participants of such tasks were often instructed to plan their strategy mentally before execution (Gilhooly et al. [2002]), what lead to better results in comparison to studies were no instructions were given. Ward and Allport [1997], as well as Unterrainer et al. [2003], showed that longer preplanning time is correlated with better performance. In addition, evidence that strategic eye movement patterns during the planning phase serve as good indicators of participants being efficient problem solvers was given by Hodgson et al. [2000].

1.2. Eye movements and the VSTM

Visual behavior is guided by image information, which is acquired and maintained by gaze shifts and the visual short-term memory (VSTM). The VSTM is a subdivision within the visual part of the working memory (Baddeley [1992] and Baddeley [2003]), which is characterized by a limited storage capacity (Phillips [1974]). There might be an upper limit on storage in terms of the total amount of information, which is characterized through a functional dependence between the total number of objects that can be stored and their complexity (Alvarez and Cavanagh [2004]). If the storage capacity is exhausted, gaze shifts can be used to gain information that is not currently represented in memory. Inamdar and Pomplun [2003] varied the distance between relevant image regions to demonstrate a trade-off between gaze shifts and VSTM storage.

Nitschke et al. [2012] assumed that regularities in cognitive processing are reflected in overtly observable eye movement patterns. Therefore, eye movements were recorded while participants worked on Tower of London (TOL) problems, in which different task demands were experimentally manipulated. Single-trial saccade-locked analysis showed that problems with higher demands on forming internal representations are associated with an increased number of gaze shifts between the presented start state and goal state, whereas no effect was found in terms of inspection durations.

1.3. Cognitive stages in mental planning

It is well known that the dorsolateral prefrontal cortex (dlPFC) plays an important role in the cortical network involved in planning and problem solving (Owen et al. [2005]). In a more recent functional magnetic resonance imaging study (fMRI), Kaller et al. [2011] used a Tower of London task to show a double dissociation between the left and right dlPFC and two cognitive aspects of problem solving. They found that a stronger activation of the left dlPFC is a sign of higher demands on the structural analysis of information, whereas stronger activation of the right dlPFC is associated with higher demands on integrating information into a sequence of actions. These findings suggest that there might be separable cognitive stages to which the left and right dlPFC contribute differently. During the first stage a mental representation is established, while during the second phase the actual planning process takes place. Earlier, based on the assumption that cognitive processing is to some extend reflected in eye movements, Hodgson et al. [2000], as well as Kaller et al. [2009] proposed that those stages might follow an at least partially sequential order with an internalization process to construct a mental problem before the planning process to solve the problem. More recent data provided by experiments like Kallers' fMRI study [2011] yielded additional support for the assumptions of a functional, as well as temporal, separation of internalization and actual planning.

1.4. Aim of this study

Previous studies showed that mental planning is a form of problem solving (Goel [2002]), which often was subject of research in the context of disk transfer tasks like the Tower of Hanoi (TOH) problem. While Unterrainer et al. [2003], as well as Ward and Allport [1997] showed that longer preplanning times during such experiments correlate with better performance, Hodgson et al. [2000] provided evidence that eye movement patterns during this planning phase also serve as good indicator for performance. It was postulated that regularities in cognitive processing are reflected in overtly observable eye movements (Nitschke et al. [2012]) and based on such an assumption Hodgson et al. [2000] and Kaller et al. [2009] proposed a temporal and functional separation of two stages of mental planning: An internalization process followed by an actual planning process.

The aim of this study was to investigate more closely the actual planning process and the importance of eye movements during that process. For this purpose, based on the TOH a disk transfer task was created, during which the participants' eye movements were recorded. To solve those TOH problems a very basic strategy was postulated and translated into respective eye movements. The assumed eye movements were compared to the participants' actual eye movements to test for the strategies' application. Therefore, the aim of this study was to show the utilization of a simple eye movement strategy during the actual mental planning process of Tower of Hanoi problems.

2. Methods

The experiment consisted of three different tasks that were conducted by 17 participants, six females and 11 males, with age varying from 21 to 28 years. All participants had normal or corrected to normal vision with only contact lenses allowed. The three tasks were solved in one sessions with short breaks of up to five minutes between each tasks. Subjects were naive to the purpose of the experiments.

2.1. Hanoi task

The Hanoi Task is the main experiment of this study. It is based on the mathematical puzzle “Towers of Hanoi” that was invented by Édouard Lucas in 1883. In the following sections the eye tracking setup used during this task as well as the task itself and the experimental procedures are described.

2.1.1. Experimental design

During each trial an image was displayed showing a puzzle similar to “Towers of Hanoi”. On the image three poles were shown, whereby two were colored in brown and the rightmost was colored in grey. In addition, three disks of different sizes and colors, a small blue one, medium sized green one and large red one, were distributed across the poles (figure 2.1.1.a). As in the original TOH, the objective of the task was to move the disks around ending with a tower, whereby the disks were sorted by size and color and were located on the grey pole on the right hand side. The goal state was a tower with the large, red disk at the bottom, followed by the medium sized, green disk in the middle and the small, blue disk on top (figure 2.1.1.b).

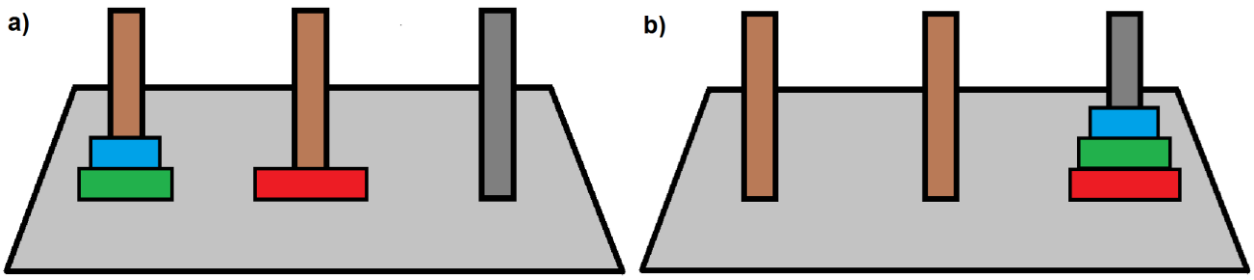


Figure 2.1.1.: An example starting state (a) of the TOH used during this study, as well as the goal state (b) used for all trials. In this version of the TOH the differently sized poles were also color coded with red for the largest disk, green for the medium sized one and blue for the smallest disk. The goal rod was also emphasized with grey color, unlike the two other rods that were colored brown. For the starting state 4-move, 6-move and 7-move problems were designed. Except for the set number of moves necessary to reach a solution these didn't follow any rules. The goal state (right) didn't vary between trials and was only explained to the participants before the experiment started. Therefore, participants had to remember this state without getting feedback later on.

The standard Tower of Hanoi shown in figure 1.1. has two internal rules: (1) only one disk can be transferred at a time; (2) a disk can only be transferred to a pole on which it will be the smallest. This TOH also has one external rule: (3) only the smallest disk on a pole can be transferred to another pole. In this case, the third rule doesn't need to be stated explicitly because the structure of the disks and poles coupled with rule 1 guarantee that it will be followed.

In this study the following rules were used:
Rule 1: Only one disk, which has to be the topmost of its pole, can be transferred at a time.
Rule 2: The disk, that is about to be transferred, can't be placed on top of a smaller disk. (i.e. in order from large to small on each pole)

Both rules were internal rules that were explicitly stated as written propositions in the instructions of the experiment and had to be memorized by the participants. To create a larger variety of trials different disk starting positions were used. The second rule had no influence on those starting disk configurations, therefore, in the beginning larger disks could be located on top of smaller disks but the participant was not allowed to recreate such configurations.

During this study, participants were only shown images of the tasks, as a consequence they were not able to solve the puzzles by moving the disks around. Instead, the participants had to solve the problems mentally and verbally report the lowest number of moves necessary to create the goal state. Supportive movements with hands or mouse as well as active speech were not allowed during this experiment. Furthermore, the participants were instructed to solve the problems as fast and precise as possible.

In total subjects had to solve 21 trials. Those were separated into three different groups of seven trials, each having its own minimal solution of four, six or seven moves. These three conditions will be referred to as complexity 4, 6 and 7 and were established by a preliminary experiment with seven participants, aged between 22 and 26 years. The order of the trials was pseudorandomized in the beginning and was the same for each participant.

2.1.2. Experimental procedure

In the beginning of the TOH task the participants received a written instruction (figure A.1.) explaining the Hanoi task. After reading the instructions the participants were asked to briefly explain the rules in their own words to ensure the correct understanding of the tasks. In case of misunderstandings or questions the participants were emended verbally. After the rules and procedure were understood the subjects' head was placed on the fixator in front of the computer screen and the eye tracking calibration started. After a successful calibration the participant had to press space to continue. Before the Hanoi image was shown, for every trial a fixation cross appeared in the middle of the screen for three seconds, which had to be focused by the subjects. After this time the cross disappeared, the trial image was shown and the eye tracking as well as the time recording started. The participant now solved the task mentally and was instructed to press the space bar as soon as a solution was found. When the space bar was pressed the image disappeared and the eye tracking, as well as time recording, stopped. Now the subject had to report verbally how many moves were necessary to build the tower and afterwards continue with the next trial by pressing space again.

After 10 trials the first part of the TOH task ended and the subject was allowed to remove their head from the fixator and take a short break. Afterwards the second part, consisting of 11 trials, was done, starting with a recalibration of the eye tracking system.

2.1.3. Eye tracking

Eye movements are subject of scientific studies since the 19th century, when the first studies were made using direct eye observations. Already in 1879 fixations and saccades were described for the first time by Louis Émile Javal, who observed the eyes of subjects while reading. These findings raised many important questions like on which words do the eyes stop and for how long which then became subject of scientific study. One of the first eye trackers was built by Edmund Huey in 1908, who connected some sort of contact lens with an aluminum pointer that moved according to the eye. In the 1970s and 1980s eye tracking research expanded quickly due to the highly discussed eye-mind hypothesis as well as the rise of computers that allowed using eye tracking results in real time.

This section describes the eye tracking setup, including the hardware and the software that was used to record eye tracking data during the Hanoi task, as well as the recording procedure will be explained.

2.1.3.1. Eye tracking setup

In this study participants had to solve tasks on a 19-inch monitor with a resolution of 1280x1024 on which a Guppy Pro camera was attached. This camera was focused on the participants' right eye and recorded its movement. The setup was connected to a Windows 7 computer with EGServer (LC Technologies, 2000-2005) software, which directly sent the data to MATLAB Software R2015a, version 8.5.0.197613 (The MathWorks Company, Natick, USA). During the task, the seated subjects' heads were placed on a fixator and they were instructed to keep it there and try to move as sparsely as possible. The fixator was placed in approximately 50 cm distance to the screen and adjusted to a height that places the subjects' eye pointing to the middle of the screen.

Eye tracking loss was indicated by the tracker by a zero value for the pupil diameter. This was resulting from eye blinks or a loss of the Purkinje reflex or of the pupil. Trials would have been excluded if the percentage of eye tracking loss exceeded 15% of all time, but this was not true for any trial.

2.1.3.2. Calibration

To guarantee a reliable recording of eye movements the eye tracking system had to be individually calibrated. The calibration tool was an integrated part of the EGServer tracking software. It showed a black screen with a yellow dot moving around the computer screen and stopping at specified locations. The subject had to follow the dot with its eyes and based on the recorded eye movement and the dot location the recorded data was be adjusted during the trials. In cases were the difference between the yellow dot position and the assumed fixation position was too large, the calibration was restarted automatically. After about half of the trials the subject was allowed to leave the fixator and take a little break, therefore the calibration had to be repeated before the start of the second part of the trials.

2.1.4. Data analysis

Data analysis was conducted using MATLAB Software R2015a, with the Psychtoolbox version 3.0.12. After preprocessing the result data, information about the duration, location and order of gazes was extracted. Gazes with a duration of at least 120ms were termed fixations. In addition, the task image was divided in three separate zones, 1, 2 and 3 along the x-axis, each containing one of the three poles. All of those zones, which are shown in figure 2.1.2., was 315 pixels wide and centered on the respective pole.

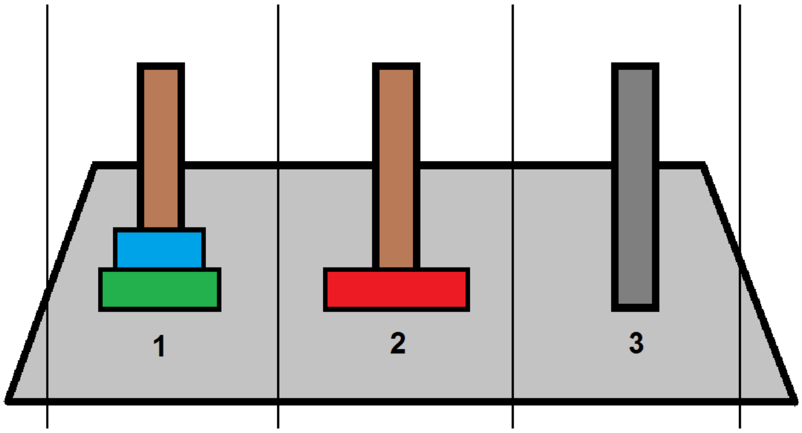


Figure 2.1.2.: The working space was separated along the x-axis in three different zones, 1 to 3. These zones were centered around the three poles and were used to classify the fixations into three categories. Consecutive fixations that occurred within the same zone were combined by averaging the x- and y-coordinates and adding up their durations up. This resulted in combined fixations, which resemble gaze shifts between the three zones.

Successive fixations that occurred in the same zone were combined by averaging the x- and y-locations and adding the durations up (figure 2.1.3.b). The order of those combined fixations, further called gaze shifts, now showed the shifts between the three zones and allowed the establishment of a gaze path (figure 2.1.3.c). This path illustrated the order, in which each participant focused on the different poles, and could be denoted as a string, which was termed working string (figure 2.1.3.d).

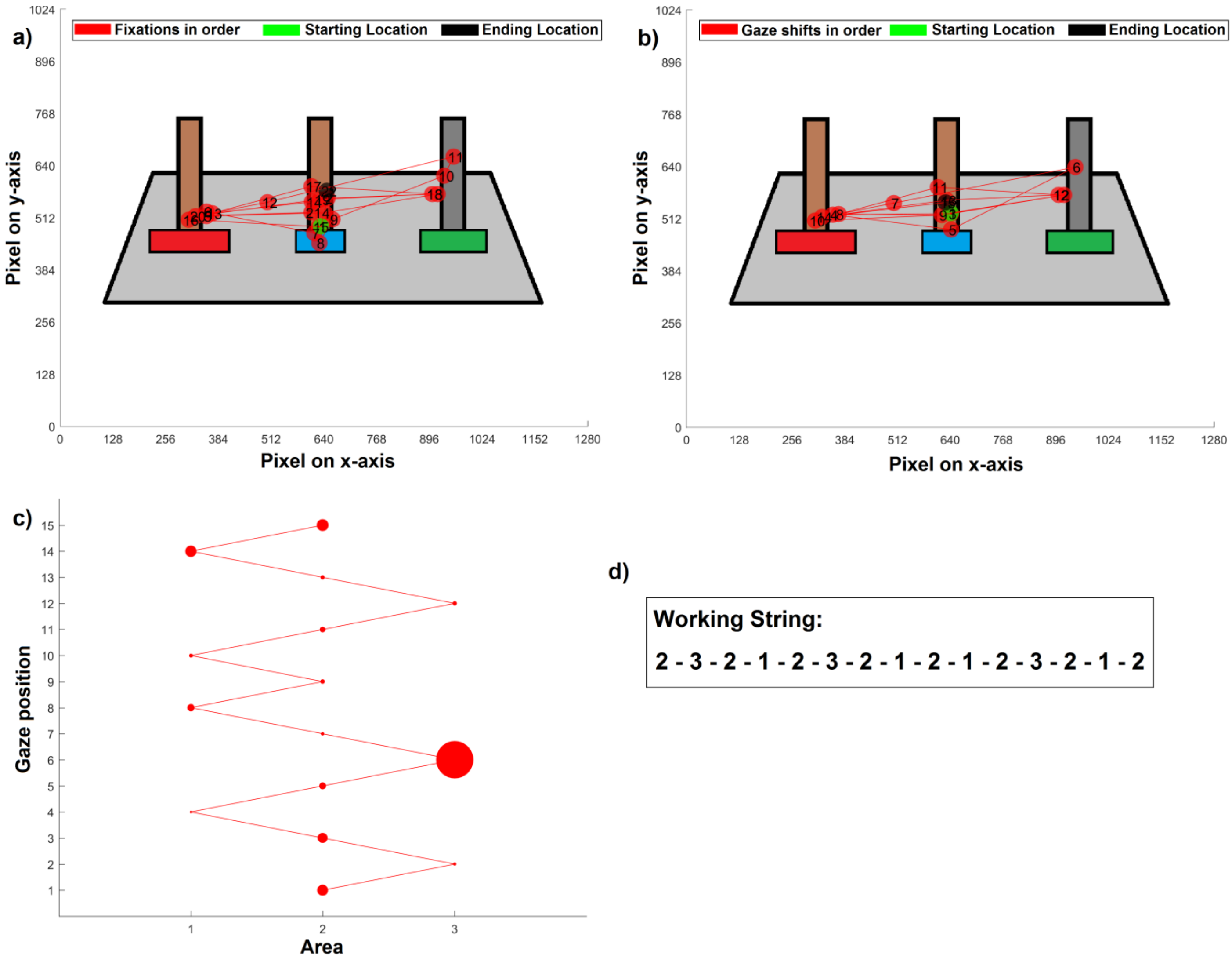


Figure 2.1.3.: Data processing from fixations (a) to combined fixations (b), the gaze path (c) and the working string. Figure a) illustrates example data of a solution to a 7-move problem. Here, the gaze data was already reduced to fixations of a duration of 120ms or longer, which are shown in order. After reducing the gazes, the resulting fixations were combined according to the three zones as well as their time of occurrence (b). The order of those fixations now revealed the gaze shifts between the zones, which could be illustrated as a gaze path (c) or as a string (d).

To analyze the working strings further, target strings were defined for each trial. For these strings it was assumed that there was a gaze shift for each picking up, as well as placement of a disk. Therefore, for trials of complexity level 4, target strings with a length of 8 were defined, for complexity level 6 the length was 12 and for level 7 it was 14.

In the following analysis the Levensthein distance was used to compare the target and working strings of each trial. This was done by scanning the working string with the target string. For this, the target string was first compared to the first segment of the working string, which length equaled the target string length. If the resulting edit distance was equal or lower than a specified threshold a match was found. This threshold was defined separately for every complexity level and is the equivalent of a random string conformity rate of less than 5%. In case of a match, the scanning position was shifted by a number of position equal to the target string length, otherwise the scanning position was shifted by one. This procedure was repeated until the end of the working string was reached, an example can be seen in figure 2.1.4.

Replacement only

Target string: **2 – 3 – 1 – 2 – 1 – 3 – 2 – 3**
Working string: **1 – 2 – 3 – 1 – 3 – 3 – 2 – 3 – 3 – 2 – 3 – 2 – 3 – 1**
Edit distance: 5x replacement (1) = **5**

Subtraction and addition

Subtraction: **1 – 2 – 3 – 1 – 3 – 3 – 2 – 3 – 3 – 2 – 3 – 2 – 3 – 1**
Addition: **2 – 3 – 1 – 3 – 1 – 3 – 2 – 3 – 3 – 2 – 3 – 2 – 3 – 1**
Target string: **2 – 3 – 1 – 2 – 1 – 3 – 2 – 3**
Edit distance: 1x subtraction (1) + 1x addition (1) + 1x replacement (1) = **3**

An edit distance of 3 is still above the 4-move threshold of 2 and the target string is moved by 1 position.

Target string: **2 – 3 – 1 – 2 – 1 – 3 – 2 – 3**
Working string: **1 – 2 – 3 – 1 – 3 – 3 – 2 – 3 – 3 – 2 – 3 – 2 – 3 – 1**
Addition: **1 – 2 – 3 – 1 – 3 – 1 – 3 – 2 – 3 – 3 – 2 – 3 – 2 – 3 – 1**
Target string: **2 – 3 – 1 – 2 – 1 – 3 – 2 – 3**
Edit distance: 1x addition (1) + 1x replacement (1) = **2**

A match was found and the target string is moved a number of steps equal to its length to scan for additional matches.

Target string: **2 – 3 – 1 – 2 – 1 – 3 – 2 – 3**
Working string: **1 – 2 – 3 – 1 – 3 – 3 – 2 – 3 – 3 – 2 – 3 – 2 – 3 – 1**
An overlap of 3 forces 3 additions and therefore a string distance of at least 3. This is above the threshold and the scanning is stopped.

Figure 2.1.4.: Example of the string scanning procedure of a 4-move problem. First the target string was compared to a number of positions of the working string equal to its length. In this case, there were five differences and therefore 5 replacements, which were each valued with a cost of 1. The resulting edit distance was 5. The Levenshtein distance also allows subtractions and additions, which were both valued with a cost of 1. Usage of these operations reduced the edit distance to 3, what was still above the threshold for 4-move problems (2). Therefore, the target string was moved along the working string by one position. On this new position only one addition and one replacement was necessary to match the strings, resulting in an edit distance of 2 and therefore a match was found. In this case, the target string was moved a number of positions equal to its length (here 8). Now three positions exceeded the working string, what resulted in the need of three additions and therefore a minimum edit distance of 3. Hence, no more matches could be found and the scanning procedure was stopped.

The thresholds were established by calculating the probabilities for random working strings to match solution strings of the same length. These calculations showed that for complexity 4, the chance of two such strings having an edit distance of 2 or less is below 5%. Since, in contrast of the Levenshtein distance, these calculations did not regard subtractions and additions of values but replacements only, additional simulations were done. Each complexity level was simulated with 1000 random generated working strings with length equal to the respective target string. These simulations confirmed the previously assumed thresholds with random match probabilities below a 5% chance. For 4-move problems an edit distance of 2 or less was determined as threshold for a match (3.67%), for 6-move problems the string distance threshold was 4 (4.11%) and for 7-move problems it was 5 (4.21%). Based on this analysis the match rate was determined, which is the number of trials with at least one match divided by the total number of trials. In addition to these match rates, the working string lengths, the position of the first matches and the error rates of each subject were the most important factors for the further analysis.

2.1.4.1. Statistics

Statistical analysis was done with IMB® SPSS® Statistics, version 23.0.0.0. For the parameters match rate, working string length, first match position as well as error rate, repeated measurement analysis of variance were conducted. The complexity levels 4, 6 and 7 served as factors. First, a Mauchly test was performed for each of these parameters, in case of significance this was followed by a Greenhouse-Geisser correction. Otherwise a repeated measurements annova was carried out. To analyze the interrelations between the presence of matches and errors during trials a contingency table was created (table A.1.) and Pearsons' chi-squared test was performed.

In addition, the data was analyzed for correlations between the parameters match rate, working string length and error rate. These correlations were performed for every complexity level as well as the combined values across all complexities. Spearman's rank correlation coefficient was used to test for those correlations. In addition, due to the low number of participants, Pearson's product-moment correlation coefficient was also calculated. An overview of the statistical test results is given in table A.2.

2.2. N-back task

The n-back task was one of the two supplementary tasks used in this experiment. It is a continuous performance task introduced by Kirchner [1958] that is commonly used as an assessment to measure part of the working memory. During an n-back task, participants have to keep a representation of recently presented items in mind, which have to be updated continuously to keep track of the current stimuli. Therefore, during this task the subject needs to maintain and manipulate working memory information.

2.2.1. Experimental design

Throughout this task, participants were presented a sequence of stimuli and had to indicate if the current stimulus matched the one shown n-steps earlier in the sequence. As part of this study two spatial n-back tasks, a 2-back as well as a 3-back task were performed. For a spatial n-back task the items, in this case letters, modalities are not important, instead the participant had to decide if the positions of the current letter matched the position of the one n-steps earlier. While in the 2-back task the item positions two turns back have to be memorized, it is three turns back for the 3-back task (figure 2.2.1.).

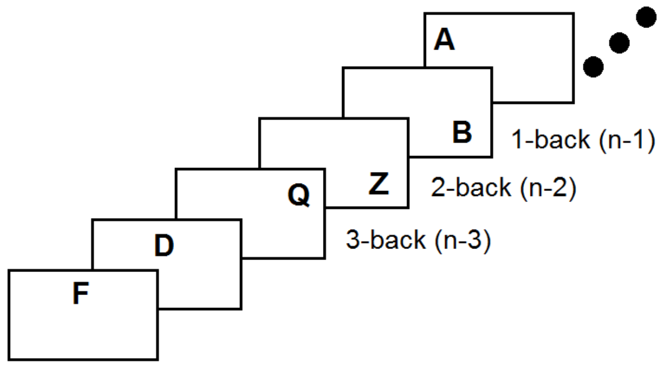


Figure 2.2.1.: Illustration of the spatial n-back task. During the spatial n-back task participants were shown a series of objects (here letters) on different screen locations. The task was to compare the object position of the current one with the one n images before and decide if the position was the same or a different one. The n-back task is a continuous task, therefore the participants were shown a set number of images in direct succession and had to remember and update information about multiple object positions and their order.

After a short explanation each participant started with a short 2-back training consisting of 10 stimuli. Participants had to either press “j” if the stimuli matched the position of the one two turns back or alternatively “f” if the position was different. During the training the letters were shown one second, followed by a three second response window, during which button input was enabled. For the following 2- and 3-back tasks, consisting of 52 and 53 stimuli and therefore 50 decisions each, the stimulus time was reduced to 0.5 seconds and the response window to two seconds. After the training and between the 2- and 3-back task participants were allowed to take short breaks.

2.2.2. Data analysis

According to the signal detection theory, D-prime values were calculated for both tasks and each participant. Furthermore, correlation analysis according to Spearman and Pearson were performed between the D-primes of the 2-back and the 3-back task as well as with the parameters of the Hanoi experiment. An overview of these correlations can be seen in tables A.3. and A.4.

2.3. Travelling salesperson task

The final supplementary experiment, a travelling salesperson task or travelling salesman problem, further denoted as TSP, was first formulated mathematically by W.R. Hamiltion and T. Kirkman in the 1800s. Combinatorial optimization problems involve the search for an optimal solution among a discrete collection of possible solutions. The TSP is a nondeterministic polynomial-hard, or NP-hard, combinatorial optimization problem in the complexity theory. It has shown that exact solutions lead to exponential time complexity and as a result, the only practical means for human solutions to this problem are of a heuristic nature. In general, human solutions of the TSP have proven to be accurate as well as rapid (Best [2006]). In this study the TSP was used as an assessment of visual problem solving capacities.

2.3.2. Experimental design

The Traveling Salesperson task consists of attempting to find the shortest tour through a series of points, starting and ending with the same points. Participants were shown 10 consecutive images, each consisting of 10 circles on a white background, seen in figure 2.3.1. Within each image the participant had to choose a starting point, by clicking on the respective circle with the mouse, and afterwards visit every other circle ending again at the starting location. This was again done by mouse clicks on the respective circle, what lead to a line appearing which connected the previous circle with the current one, so the participants were able to see the route they already took. After ending on the starting circle the image of the current trial disappeared and the next trial began.

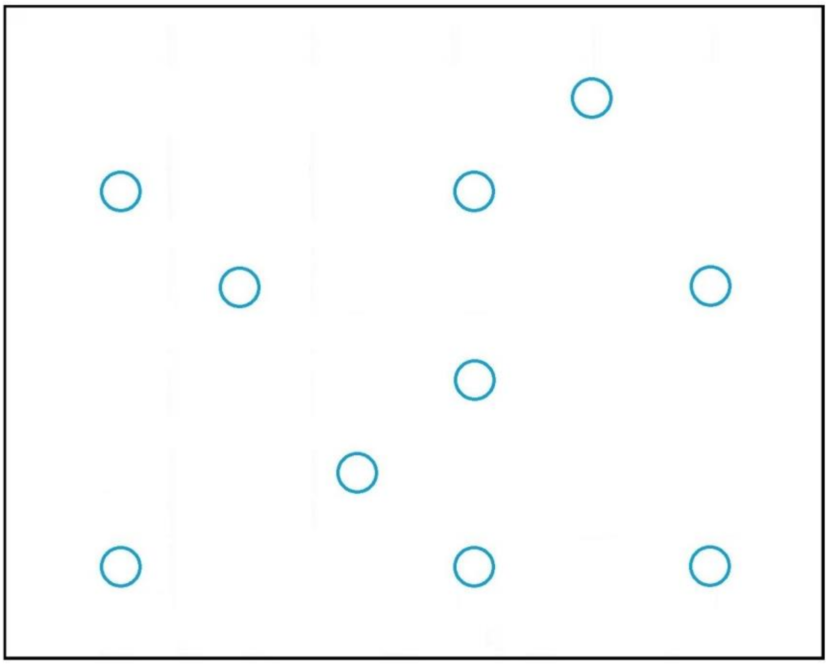


Figure 2.3.1.: Example image of the travelling salesperson task used in this study. Participants had to connect all circles while starting and ending on the same one, which they could choose freely. They were instructed to connect the circles in a way, that the resulting route was as short as possible.

2.3.2. Data analysis

For every trial the optimal route was determined and the proportions of each participants’ routes to these optimal routes was calculated. From these proportions the average detour of each participant in relation to the optimal route was extracted. As in the previous experiments, these detours were used for correlation analysis according to Spearman and Pearson with the n-back and the Hanoi parameters. These results can be seen in detail in tables A.3. and A.4.

3. Results

3.1. Hanoi experiment

3.1.1. Trial complexity and error rates

During the Hanoi task three groups of trials were defined according to the number of steps necessary for the optimal solution. Based on a short preliminary experiment these groups were described as complexity levels. Looking at the error rates in figure 3.1.1. this assumption could be confirmed, because the error rate of complexity 4 was significantly lower than of complexity 6 which again was significantly lower than the rate of complexity 7 ($F(2.32) = 17.279$; $p < 0.001$; $\eta^2p = 0.519$) . The figure also shows that for complexity 6 and 7 participants often underestimated the number of necessary steps, whereas both of the errors occurring during complexity 4 were overestimations.

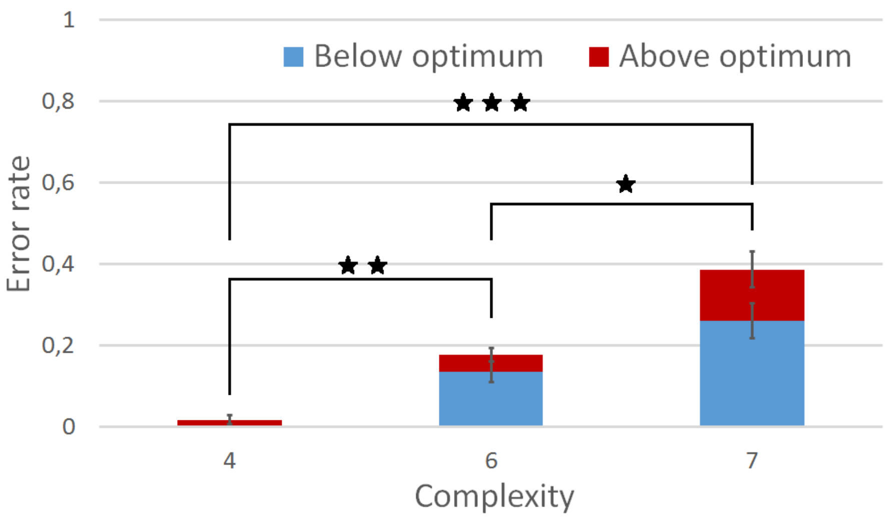


Figure 3.1.1.: Error rates of the Tower of Hanoi experiment depending on the task complexity. The number of wrong answers increased significantly ($F(2.32) = 17.279$; $p < 0.001$; $\eta^2p = 0.519$) according to complexity, with the lowest amount for 4-move problems and the highest for 7-move problems. While both wrong answers given during 4-move problems were overestimations, there were more underestimations for 6- and 7-move problems.

Looking at the error rates of every participant, which are shown in figure 3.1.2., a high variance between subjects could be found. While subject 8 and 11 did not give any wrong answer, subject 4 gave a wrong answer during every trial of complexity 7. Also there were only two participants that made one error each during complexity 4.

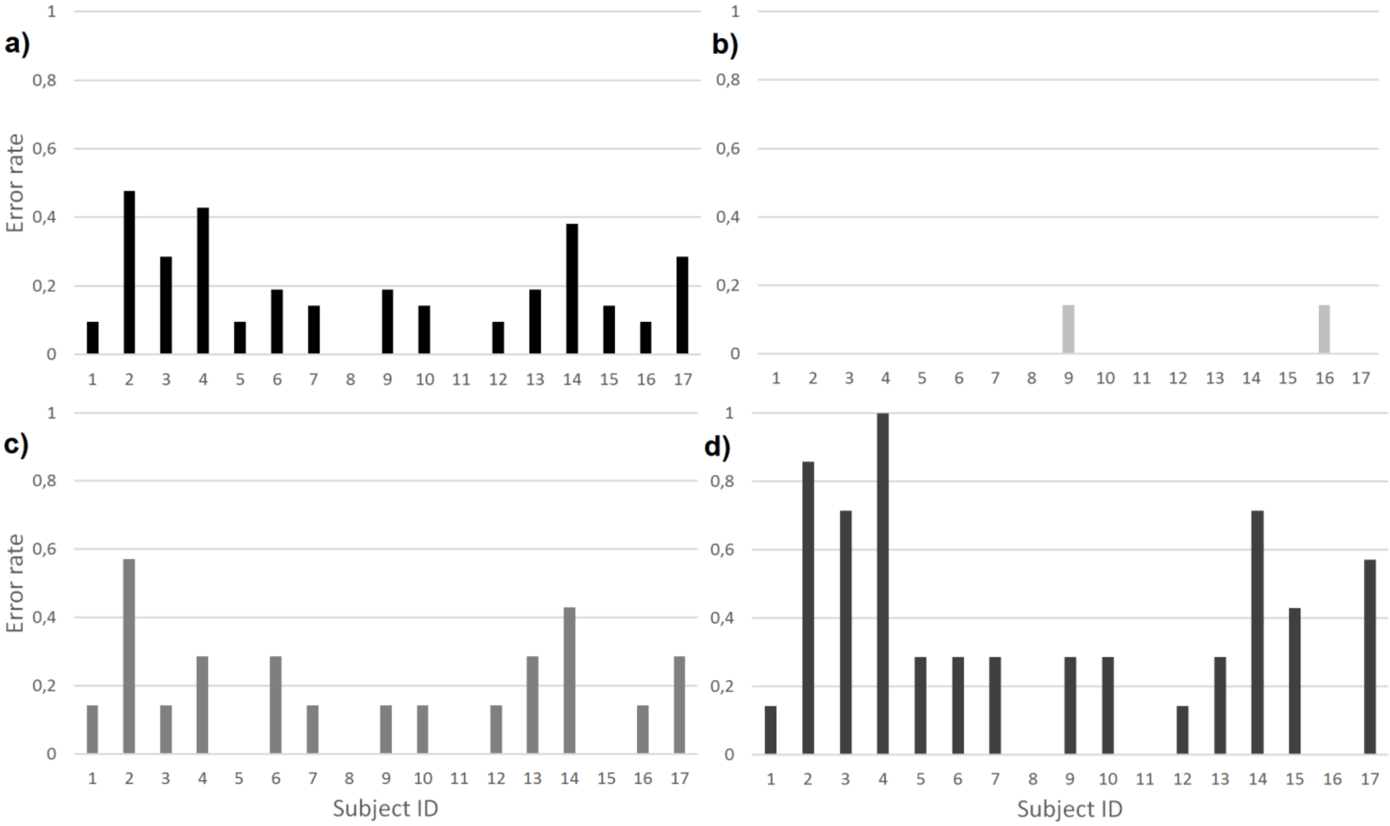


Figure 3.1.2.: Tower of Hanoi error rates across participants and complexities. The average data across all n-move problems is shown in figure a), which shows a variance of wrong responses of 0% to 47%. Looking only at the 4-move data (b), it can be seen that only two wrong answers were given, each by a different subject. For 6-move problems (c), the error rates were significantly higher with values up to 57% and only four people without errors. The error rates were highest during 7-move problems (c), here only three participants gave the right answer for every trial while one participant didn't answer correctly at all.

Since every participant solved their trials in the same order, a potential learning effect was investigated. Therefore, an analysis of the error rate over the trial order was done, which is shown in figure 3.1.3. The figure shows a small tendency for complexity 7, but no effect for the other difficulties.

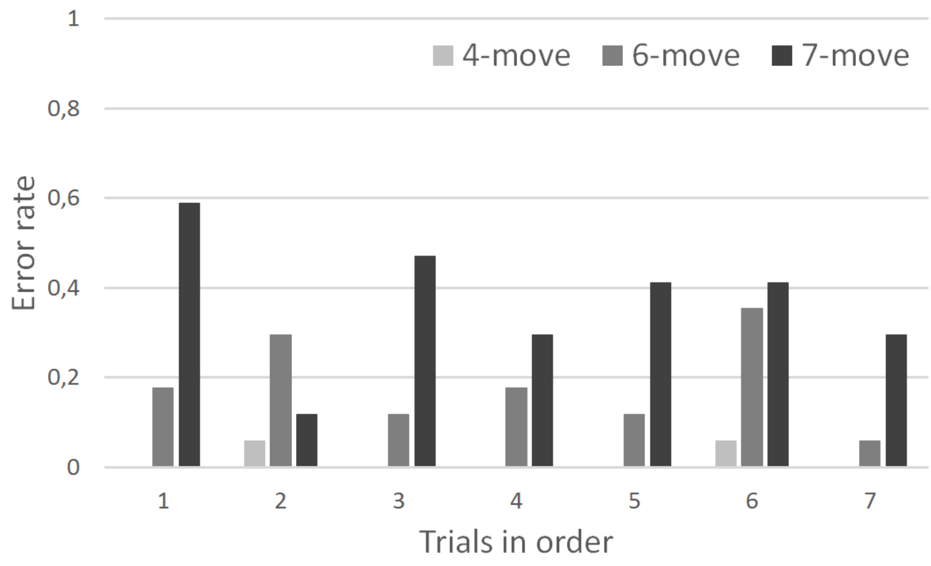


Figure 3.1.3.: Error rates for every trial, separated by complexity and in order of appearance. While the results for 4- and 6-move problems didn't improve according to the number of problems of the same complexity already solved, there was a weak tendency within 7-move problems. While repetition might improve the participants' performance of more complex tasks slightly, no clear effect was found during this study.

3.1.2. Match rate

As explained in chapter 2.1.4., the Levensthein distance was used to scan the working strings of the participants with the target string of the assumed strategy to find matches. Figure 3.1.4. illustrates the percentage of trials in which at least one match was found; these values were denoted as match rate. First of all, the figure shows that matches were found for every complexity and that this match rate varied between 24.4% with a standard error of 3.5% for complexity 4 and 47% (SE 7.03%) for complexity 7 with a total average match rate of 37.8%. While there was a significant difference in the match rate between complexity 4 and 6, as well as between 4 and 7, the match rates of complexity 6 and 7 showed almost no difference ($F(2.32) = 9.527$; $p < 0.001$; $\eta^2 p = 0.373$).

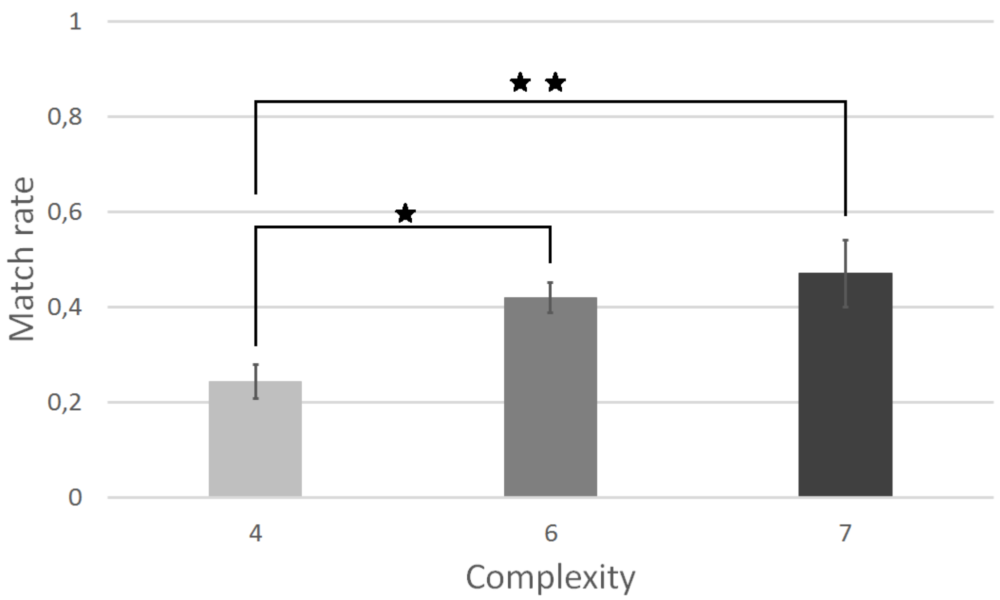


Figure 3.1.4.: Proportion of trials with matches for each complexity. While there were overall more trials without any match the figure shows that in, depending on complexity, 24% to 47% of trials at least one match was found. In addition, the 4-move problems showed a significantly lower percentage of trials in comparison to 6-move and 7-move problems ($F(2.32) = 9.527$; $p < 0.001$; $\eta^2p = 0.373$).

There were also trials with multiple matches, which are matches that occurred after another one but still within the same trial. While there were almost no multiple matches for complexity 4 and 6, the number of multiple matches within complexity 7 was significantly higher at 17.9% of the trials ($F(2.32) = 9.503$; $p < 0.01$; $\eta^2p = 0.373$). Overall the percentage of trials with more than one match was below 15%.

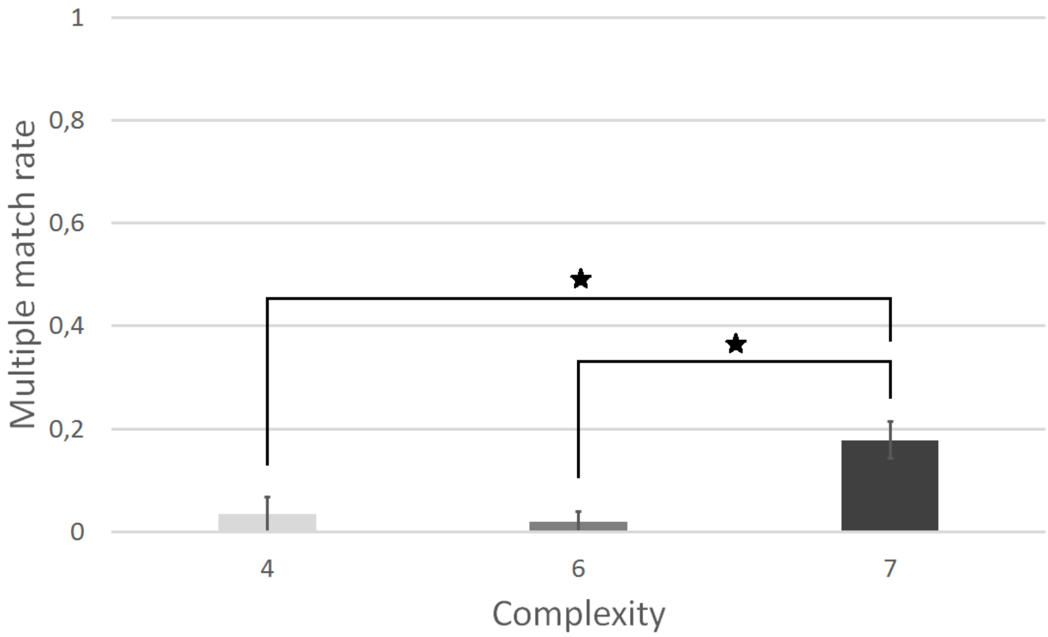


Figure 3.1.5.: Proportion of trials with multiple matches for each complexity. While only 3.45% of 4-move and 2% of the 6-move problem trials contained multiple matches, the rate of 7-move problem trials with multiple matches was significantly higher with 17.86% ($F(2.32) = 9.503$; $p < 0.01$; $\eta^2p = 0.373$).

The distribution of matches in regards to participants giving correct or incorrect answers can be seen in figure 3.1.6. It shows the percentages of matches across all subjects, as well as the errors that were made. Additionally, a high variance between subjects can be seen: While subjects like number 5 and 11 had very high match rates, subjects like number 3 and 4 only showed very low numbers of matches. Figure 3.1.6.a illustrates the averages across all complexity levels. Here it is shown that there were overall more errors during trials with no matches compared to trials with at least one match. While the amount of errors for complexity 4 (figure 3.1.6.b) was too low for analysis, the data of complexity 6 (figure 3.1.6.c) and complexity 7 (figure 3.1.6.d) followed this trend, which is strongest for complexity 7.

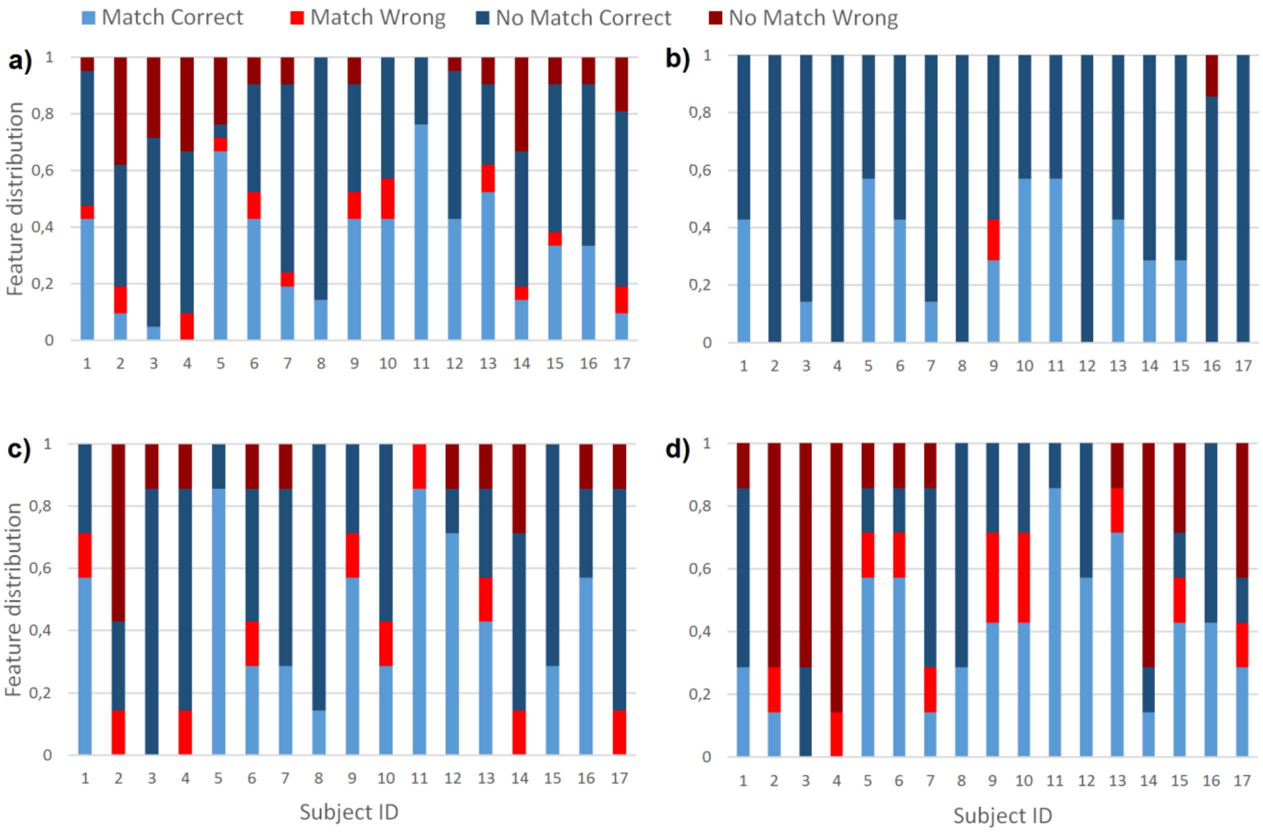


Figure 3.1.6.: Distribution of trials with errors and matches for each participant. Overall (a) there were more trials without matches and wrong answered trials without matches, than trials with matches and wrong answered trials with matches. There was also a great variance between participants concerning the error and match rates. Looking only at 4-move problems (b) it can be seen that only 2 wrong answers were given, one in a trial with a match and one in a trial without. In addition, there were some participants with matches in almost 60% of their trials while others produced no match at all. The 6-move problems (c) showed an increased error rate for most participants, only 3 gave the correct answer for every trial (5, 8, 15). While for two of those participants most trials didn't contain a match, the data of participant 5 showed matches in all but one trial. The highest error rates were found for trials with 7-move problems (d), where 4 participants showed error rates above 50% and participant 4 didn't answer a single trial correctly. Additionally, for most participants the highest amount of trials with matches was found for 7-move problems.

Correlations between the error rates and match rates across all difficulties can be seen in figure 3.1.7. Here a significant negative correlation (Pearson: $R = -0.568$; $p < 0.05$) between match rate and error rate was found, what illustrates that on average subjects with higher match rates had lower error rates. Looking individually at the different complexities, the only significant correlation was found for complexity 7 ($R = -0.552$; $p < 0.05$).

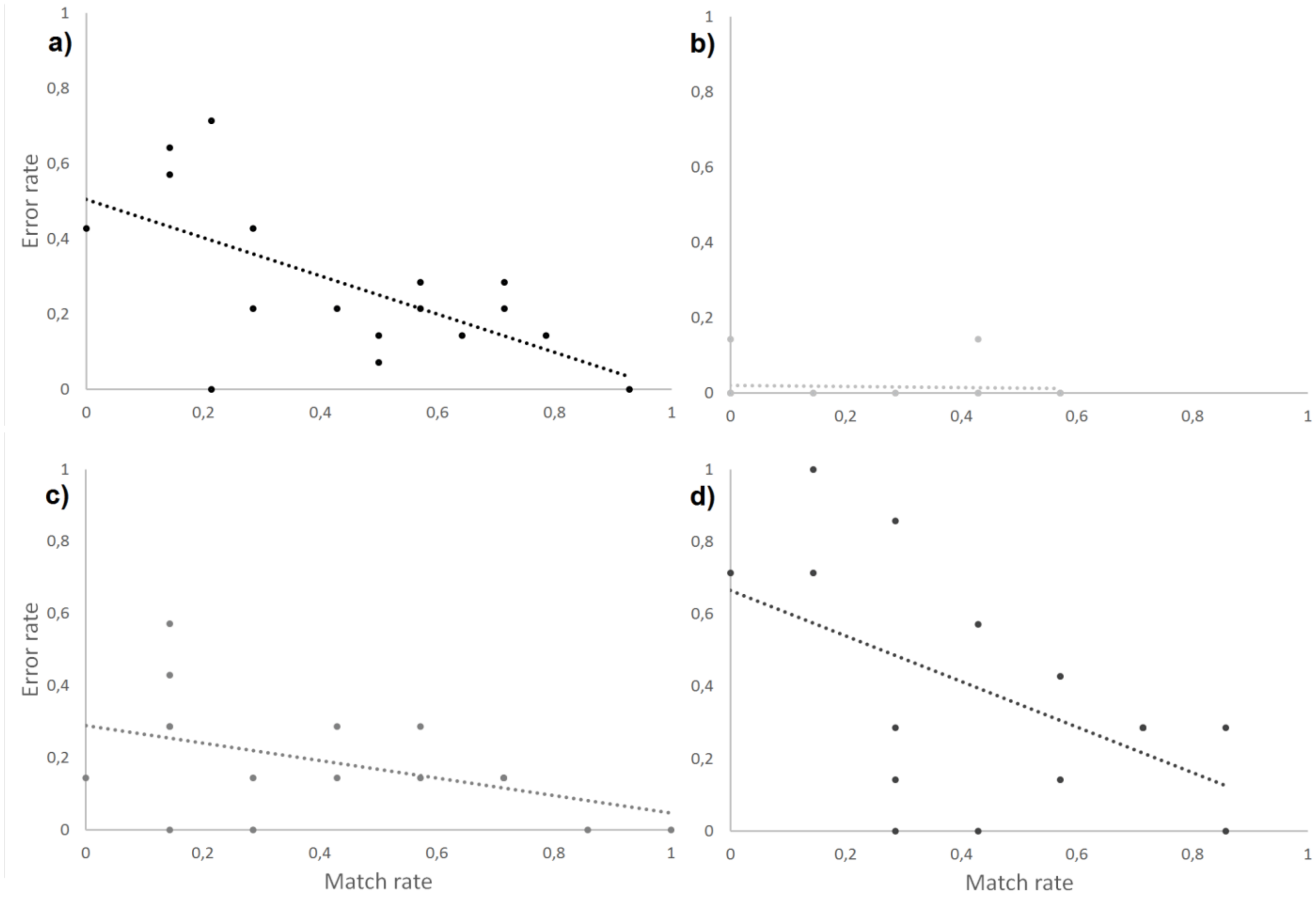


Figure 3.1.7.: Correlations between error rates and match rates of the Tower of Hanoi task according to complexity. While the only significant correlation (R = -0.552) was found for 7-move problems (d), there was a strong tendency (R = -0.460) for 6-move problems (c) and the combined 6- and 7-move trials (a). Since there were only two errors during 4-move problems (b) the data was neglected for the combined data (a).

To further analyze these findings, contingency tables with the presence of matches and errors as variables were created (table A.1.). These relations were further evaluated, using Pearson's chi-squared test. The results show that, across all complexities, there was a strong tendency of trials with at least one match being answered correctly more often, than trials were no match occurred ($p = 0.061$; $chi^2 = 3.507$; $df = 1$). Since no effect was found for the data of complexities 4 and 6, this is due to complexity 7, which showed a highly significant effect ($p < 0.01$; $chi^2 = 9.913$; $df = 1$).

3.1.3. First match position

After establishing that at least one match occurred in approximately 37% of trials, the positions of these first matches within each of those trials was examined. Since this position equals the number of areas the participant looked at until the match occurred, the starting position of the first match can be subtracted by one to obtain the number of gaze shifts between the three areas. Figure 3.1.8. illustrates that, although the average position of complexity 6 was the highest, only the average first match positions for complexity 4 differ significantly from complexity 7 ($F(2.18) = 2.705$; $p = 0.94$; $\eta^2p = 0.231$).

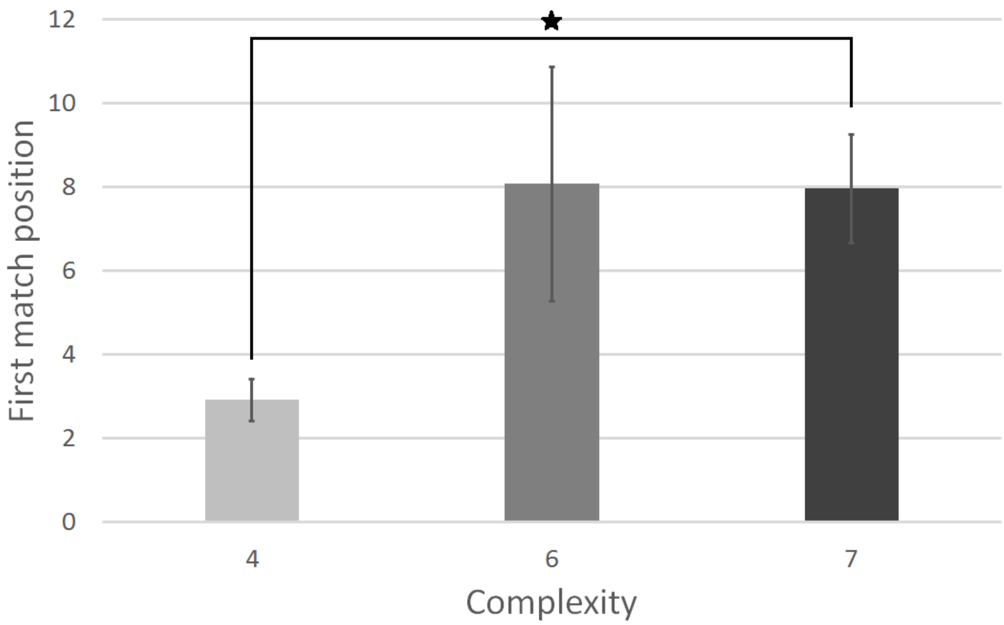


Figure 3.1.8.: Average position of the first match for each complexity. The average number of gaze shifts between the three zones of the working area before the start of the first match was significantly higher for 7-move problems than for 4-move problems. While the average position of 7.964 for 7-move problems was lower than the 8.079 of 6-move problems, there was no significant difference between 4- and 6-move problems due to the high 6-move standard error of 2.795.

Looking at the data of the individual subjects, which is shown in figure 3.1.9., a great variance between as well as within subjects can be seen. The average first match position varied from 1.5 with a standard error of 1.5 for the two matches of subject 4 to an average of 32.167 with a standard error of 9.411 for subject 17. Subject 3 was the only subject with only one match, which showed a starting position of 0. This variance between subjects could again be an indicator for the use of different strategies, this time in the pre-match phase.

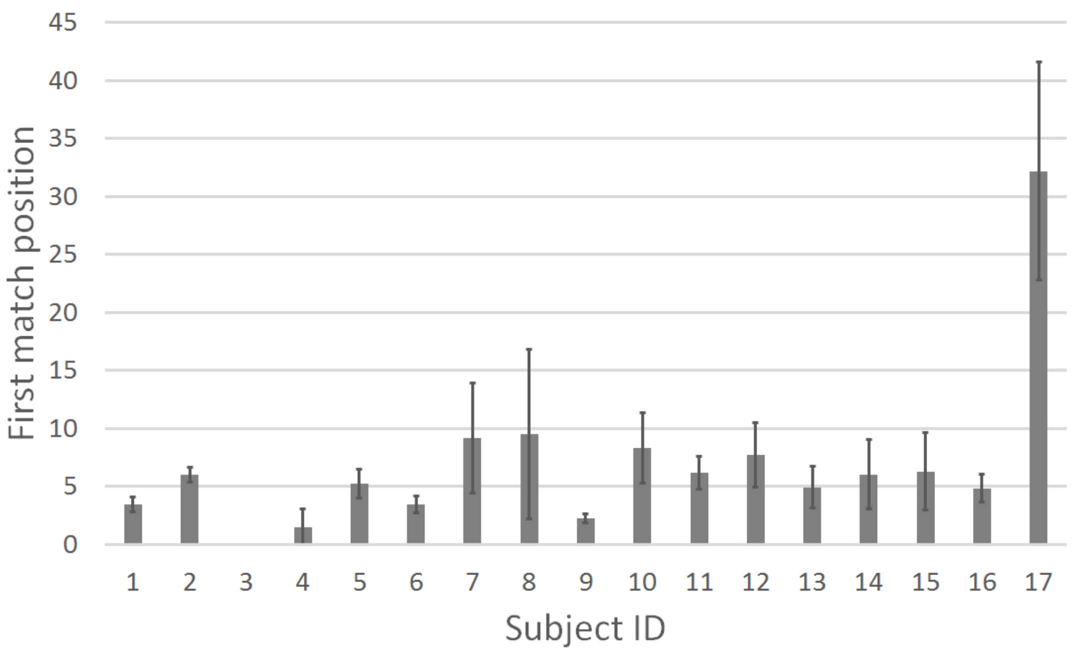


Figure 3.1.9.: Average position of the first match for each participant. A considerable variance in the average first match position could be found across all participants, as well as within some of them. The variance across all participants reached from an average starting positon of 0 for the only match of subject 3, up to 32.176 with standard errors of up to 9.411 for subject 17.

3.1.4. Working string length

The working string length is the last factor that was investigated for the Hanoi task. It equals the number of gaze shifts a participant performed between the three specified zones during a single trial. Further, by normalizing this working string length with the optimal solution of the respective complexity the string length ratio was received. The normalized string length ratio for each subject can be seen in figure 3.1.10., it varied between 0.796 (SE 0.151) for subject 3 and 2.111 (SE 0.239) for subject 10. This indicates that the average working string length of subject 3, who had only 1 match, on average was lower than the respective target string length, while the working string length for subject 10, with an average match rate of 57.14%, was more than double the target length.

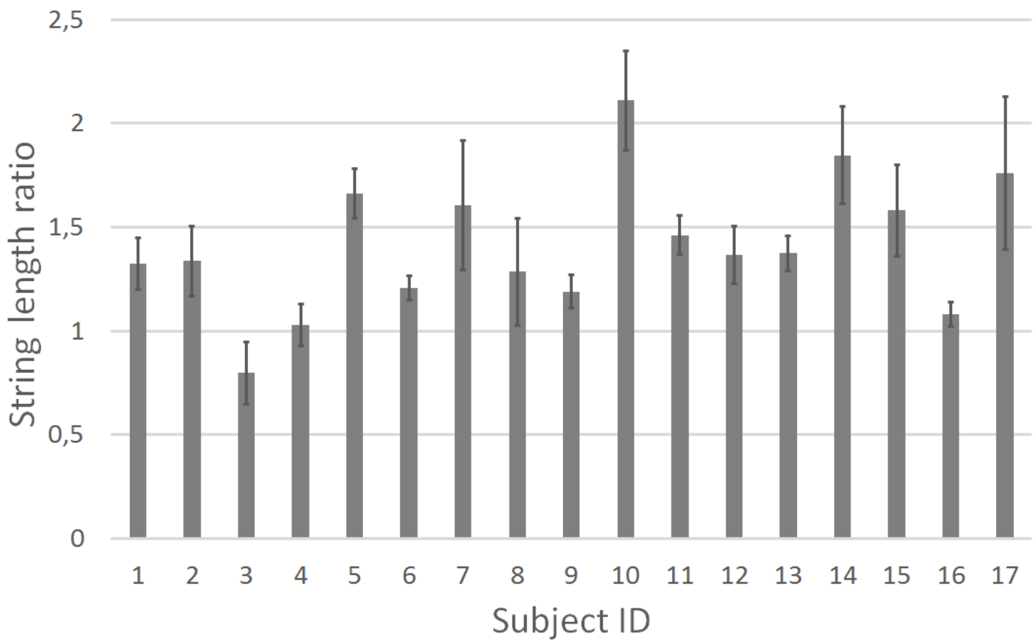


Figure 3.1.10.: Normalized string length for each participant. The string length ratio is the average length of the working strings normalized by the respective target string length. The figure shows that, while participant 3 on average didn't perform enough gaze shifts to even have the possibility of a match with edit distance 0, participant 10 on average performed more than twice the number of gaze shifts between zones than was necessary. Overall there was also a considerable variance between participants.

The string length ratio was also complexity dependent with a significant difference between the average string lengths of all complexities ($F(2.32) = 34.888$; $p < 0.001$; $\eta^2p = 0.686$). This, as well as the match rate across all difficulties is illustrated in figure 3.1.11. While the string length ratio as well as the match rate seemed to depend on the complexity level, the increase in string length ratio towards complexity 7 and the variances of the string length ratios were higher than the increase of the match rates.

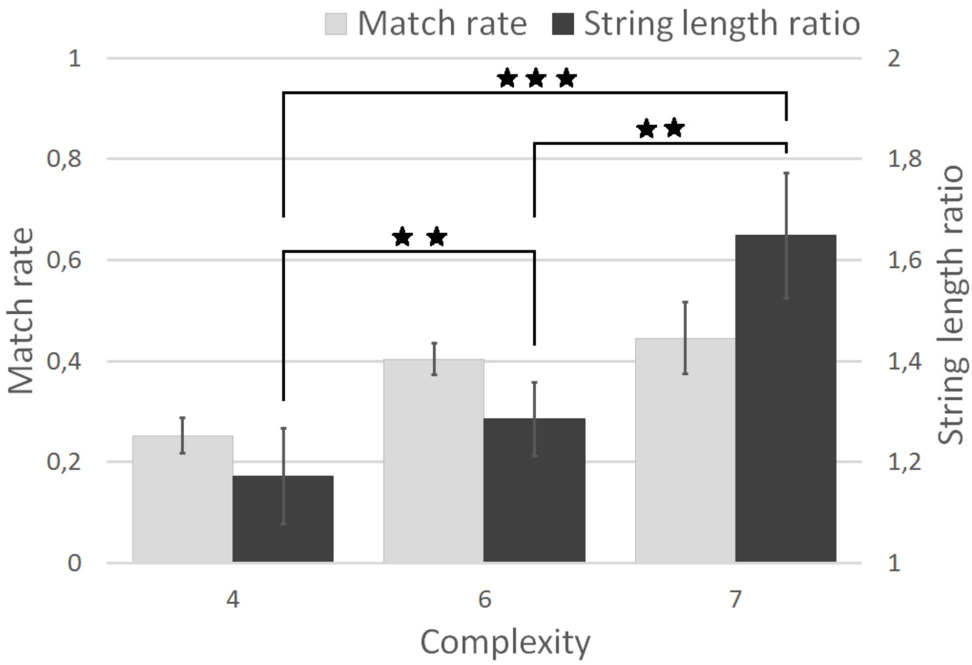


Figure 3.1.11.: Match rate and string length ratio for each complexity. While there was only a significant match rate increase from 4-move problems to 6- and 7-move problems, the string length ratio increased significantly with each increase of complexity ($F(2.32) = 34.888$; $p < 0.001$; $\eta^2p = 0.686$). Especially the increase of the string length between 6- and 7-move problems was considerable higher than the respective increase of the match rates.

Since both the string length ratio and the match rate seemed to be complexity dependent correlation analysis for every complexity was performed. Figure 3.1.12. shows these correlations, which were only significant for complexity 4 (Spearman: $R = 0.557$; $p < 0.05$), but showed trends for complexity 7 (Pearson: $R = 0.324$) and across all complexities (Spearman: $R = 0.350$). The likelihood of a match occurring therefore seemed to be higher in trials were the subjects working string length was longer.

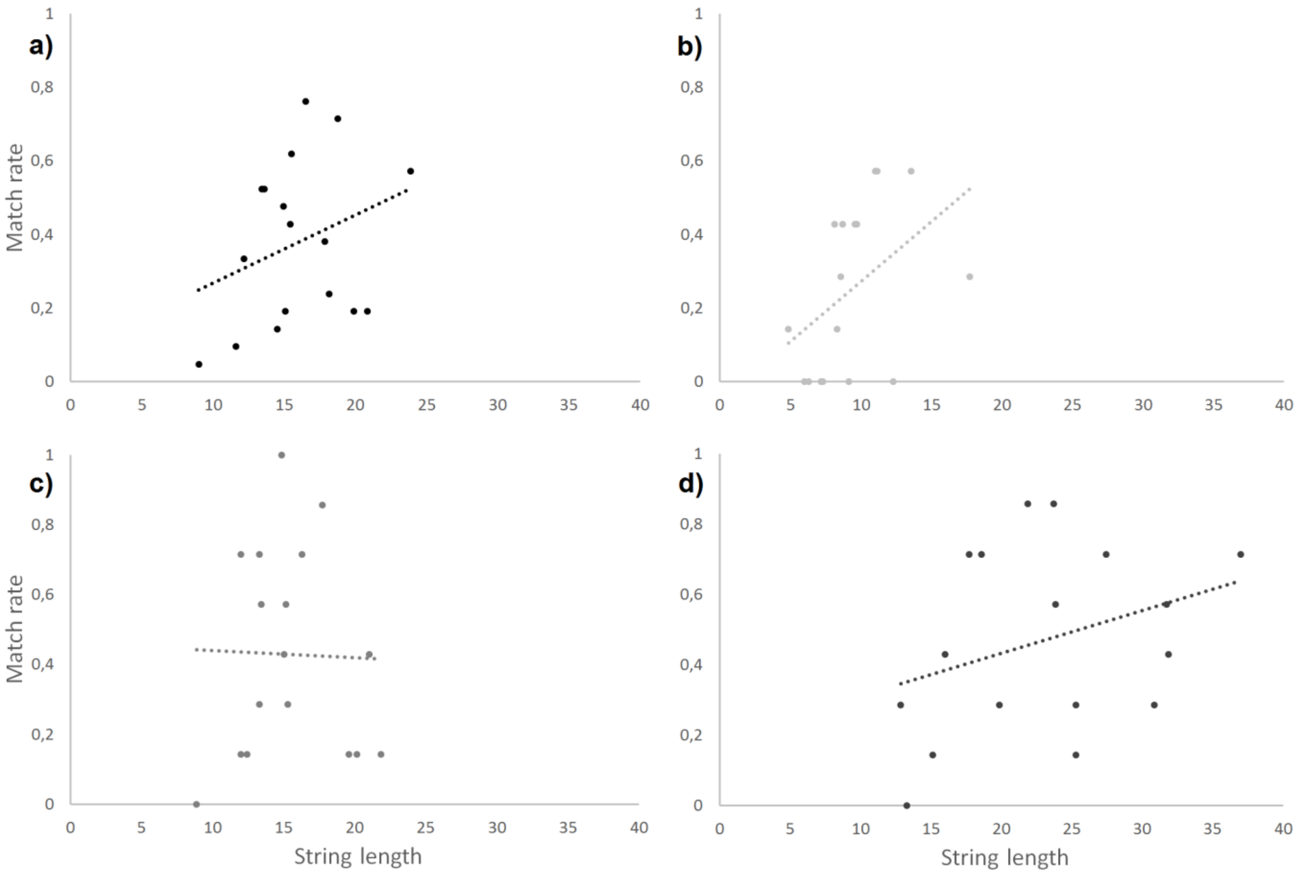


Figure 3.1.12.: Correlations between match rates and string lengths of the Tower of Hanoi task. The only significant correlation (Spearman: R = 0.557) was found for 4-move problems (b). This correlation as well as a trend (Pearson: R = 0.324) for 7-move problems (d) resulted in a trend (Spearman: R = 0.350) across all complexities (a). For 6-move problems (c) no correlation could be established.

Since it was established earlier that the error rate was also complexity dependent and a correlation between match and error rate was likely to occur, correlations between string length and error rate were also expected. Figure 3.1.13. illustrates those correlations, which showed an opposite effect for complexity 6 and 7. While for complexity 7 the error rate decreased with larger string lengths, it increased for complexity 6. Because of the low error rate during complexity 4 no correlation could be found there.

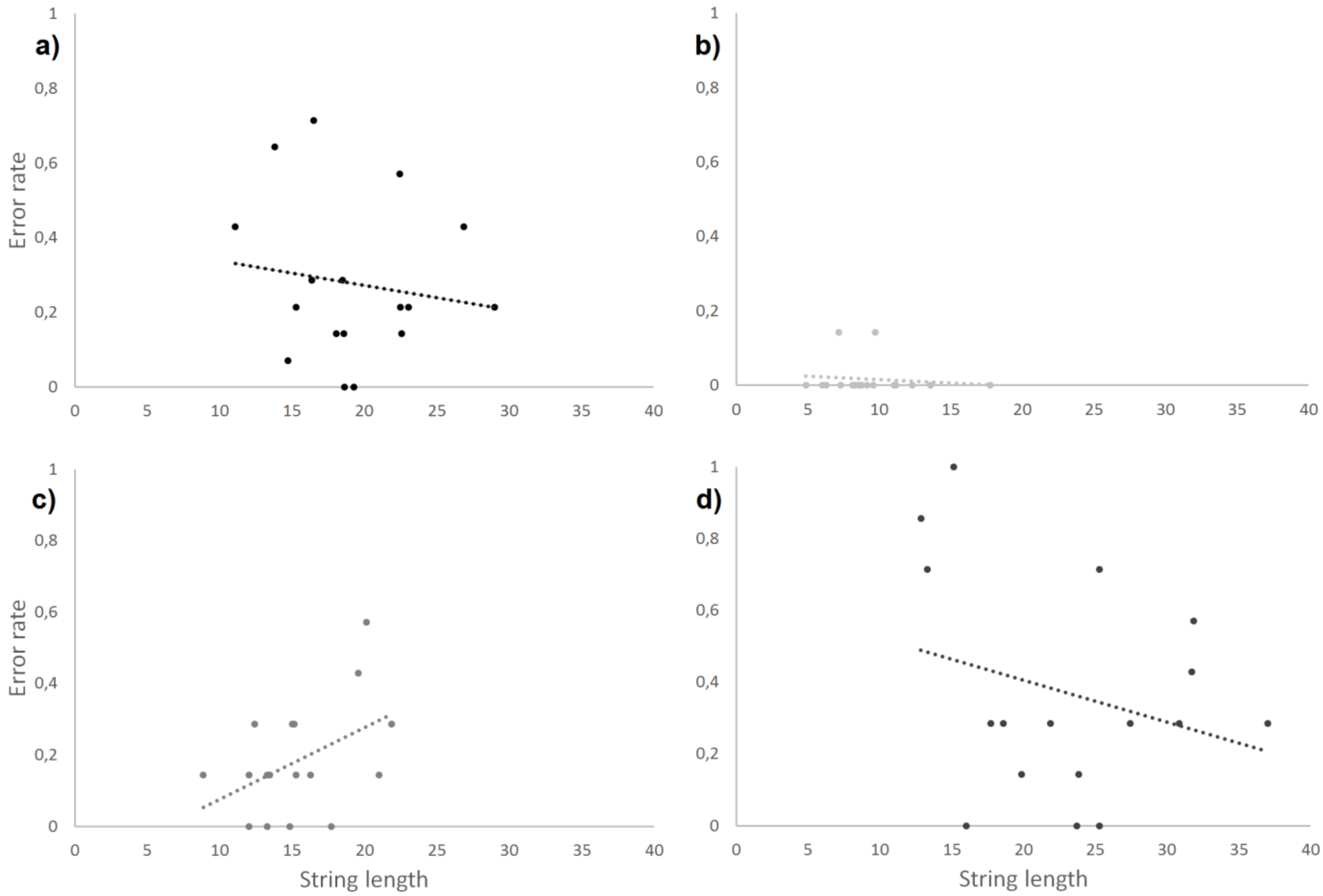


Figure 3.1.13.: Correlations between error rates and string lengths of the Tower of Hanoi task. While the strongest correlation (Pearson: R = 0.457) was found for 6-move problems (c), the analysis for 7-move problems (d) yielded opposite results (Pearson: R = -0.274). Since there were only two errors during 4-move problems (b), this data was neglected for the analysis of the overall data (a), which shows almost no effect.

3.2. N-back task

As explained in chapter 2.2., the n-back task was used as an assessment of the working memory capabilities of the respective participant. The results were analyzed based on the signal detection theory. While the average 2-back D-primes of all subjects can be seen in figure 3.2.1.a, the 3-back data is shown in figure 3.2.1.b. On average the D-primes of the 3-back task were higher than the ones of the 2-back, what was expected because 3-back tasks are considered more difficult than 2-back tasks.

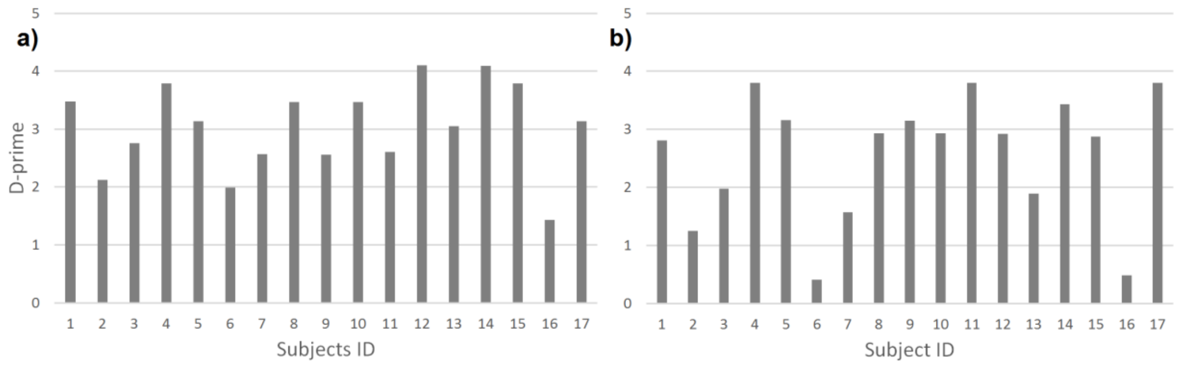


Figure 3.2.1.: D-primes of the n-back tasks for each subject. As expected, the D-primes of the 2-back task (a) were on average higher than the D-primes of the 3-back task (b), which is assumed to be more difficult. The 2-back results for most participants were slightly better than the respective 3-back results and only a few improved during the second task.

The correlation between the sensitivity indices of the 2-back and 3-back task is illustrated in figure 3.2.2. As expected, a significant correlation (Pearson: $R = 0.731$; $p < 0.001$; Spearman: $R = 0.543$; $p < 0.05$) was found, what confirmed the successful implementation of the two tasks.

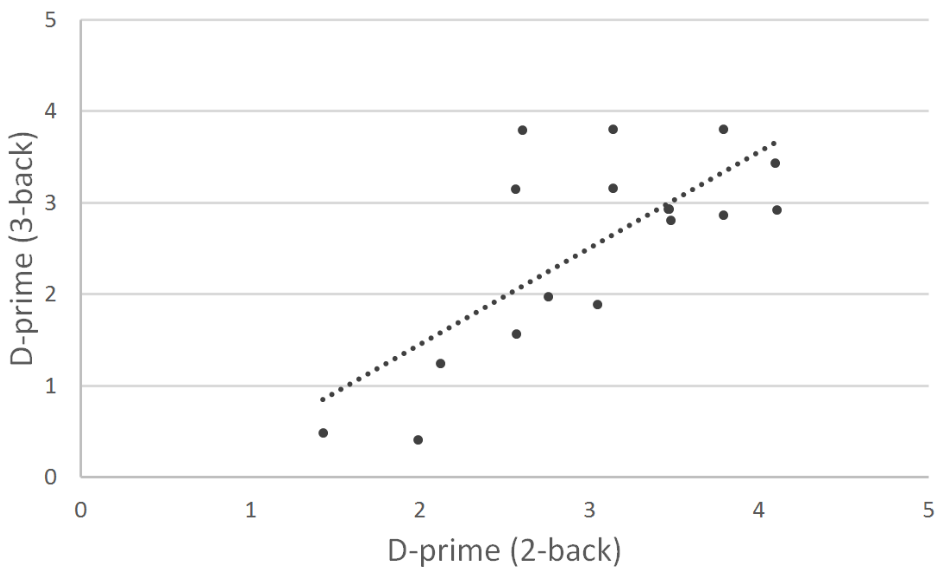


Figure 3.2.2.: Correlation between the 2- and 3-back D-primes. The data showed a significant correlation (Pearson: $R = 0.731$; $p < 0.001$; Spearman: $R = 0.543$; $p < 0.05$) between the 2-back and 3-back data. Therefore, participants with higher D-primes during the 2-back task also showed higher D-primes during the 3-back task.

3.2.1. N-back and Hanoi task

In regards to the TOH data, no significant correlations were found between the n-back values and the match rate or the working string length. Here the strongest non-significant correlations were found for the string length of complexity 7 (2-back: Pearson: R = 0.438; 3-back: Pearson: R = 0.405) as well as overall (2-back: Pearson: R = 0.350; 3-back: Pearson: R = 0.364), these correlations with the 3-back D-primes are shown in figure 3.2.3. The figure shows a tendency of participants with better n-back results making more gaze shifts during more complex trials of the TOH tasks.

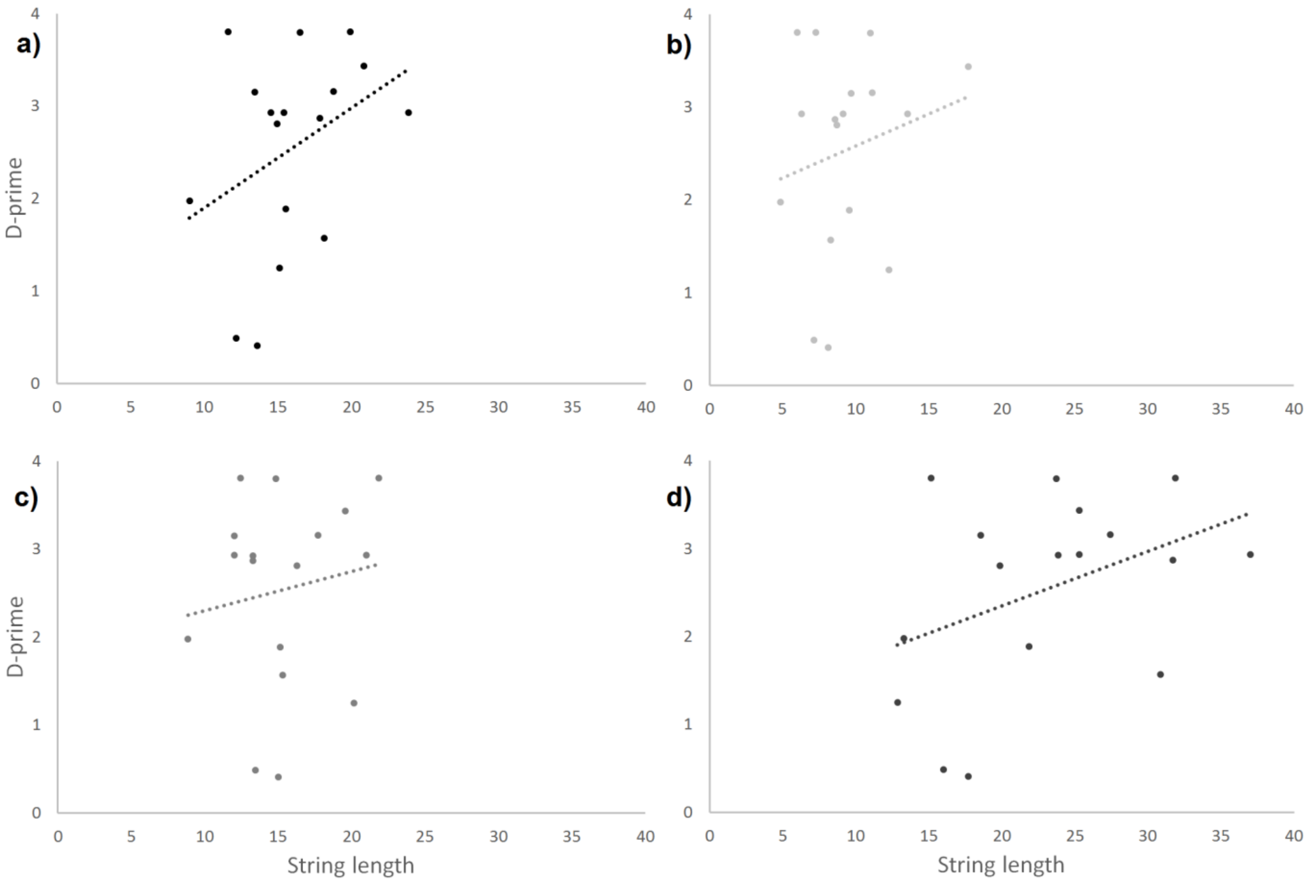


Figure 3.2.3.: Correlations between the 3-back D-primes and the Tower of Hanoi string lengths. While none of these correlations was significant, strong trends for the 7-move (d) data (Pearson: R = 0.405) as well as the overall (a) data (Pearson: R = 0.364) could be found. The 4- (b) and 6-move (c) problems didn't show enough variance for distinct trends.

Correlation analysis between the n-back data and the TOH error rates only yielded significant results between the 2-back D-primes and the number of errors during complexity 4 of the TOH experiment (Pearson: $R = -0.515$; $p < 0.05$; Spearman: $R = -0.485$; $p = 0.048$). Since only two participants made one error each during complexity 4 of TOH and no other correlations or trends could be identified between n-back data and TOH error rates, this correlation has to be treated with some reservation.

3.3. Travelling salesperson task

The TSP was used as an assessment of visual problem solving capacities. For data analysis the length of the participant's route was divided by the length of the optimal route and the average detour was calculated in percent. The average detour for each participant can be seen in figure 3.3.1. It is illustrated, that the average detour for each participant was below 5%, what confirms Bests' [2006] findings, of human TSP solutions being very accurate. In addition, there is a high variance between subjects, reaching from a 0% detour for subject 6 to a 4% detour for subject 17.

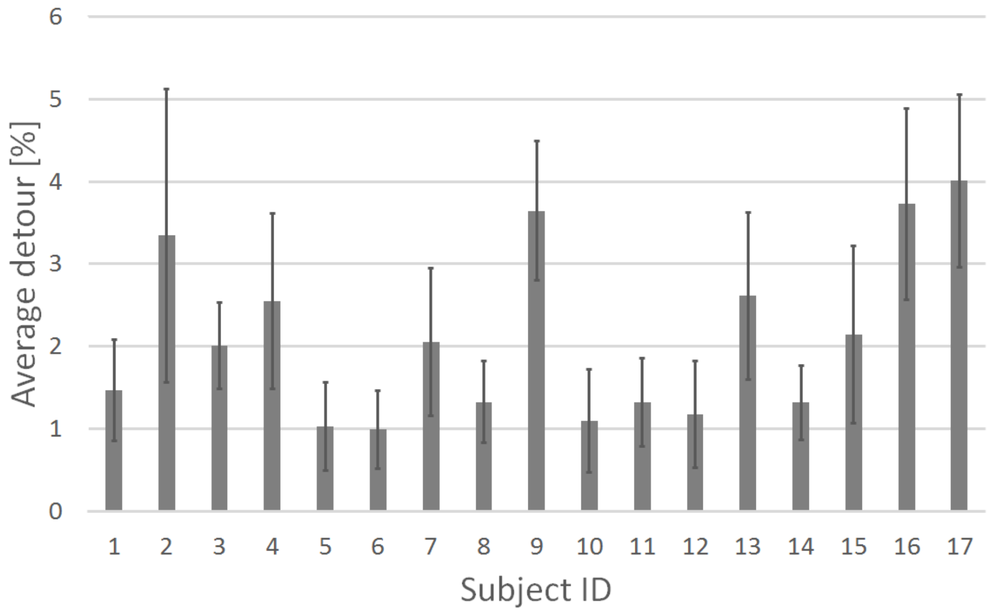


Figure 3.3.1.: Average detour of the travelling salesperson task for each subject. The average detour is the average additional distance the participant chose during the travelling salesperson task, expressed in a percentage of the optimal route. Hence, an average detour of 5 denotes, that the participant on average chose routes which are 5% longer than the respective optimal routes. While the overall results were very accurate with an error margin below 5% of the optimal routes, there were notable variances between participants with overshoots from about 1% to 4% of the optimal routes.

3.3.1. TSP and Hanoi task

Looking at the match rate of the Hanoi experiment, a significant correlation between the 4-move match rates and the TSP detour could be found (Spearman: $R = -0.504$; $p < 0.05$), which is shown in figure 3.3.2. The figure illustrates that participants with a lower average detour in the TSP task tended to have more trials with matches during complexity 4 of the Hanoi task.

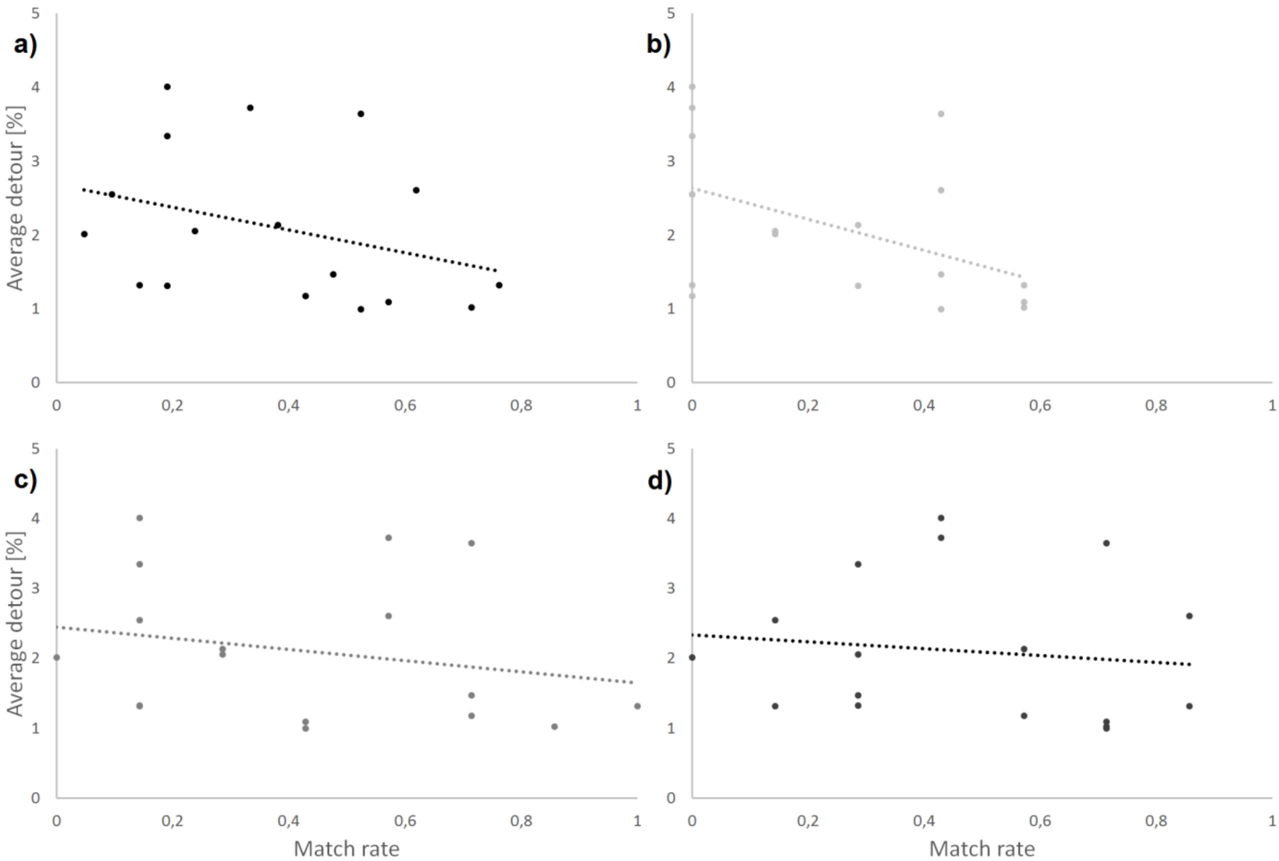


Figure 3.3.2.: Correlations between the average detours of the travelling salesperson task and the match rates of the Tower of Hanoi task. The figure shows a significant correlation for the 4-back (b) data (Spearman: R = -0.504; $p < 0.05$) as well as a trend for the overall (a) data (Spearman: R = -0.361). Both of these correlations were negative, therefore participants with better TSP results showed higher match rates during the 4-move TOH. No correlations could be established for the 6-move (c) and 7-move (d) data.

The correlation analysis for the average detours with the Hanoi working string lengths didn't yield any significant results, but negative trends could be found. While the strongest tendency was found for 4-move problems (Spearman: R = -0.346), the tendencies for the overall string length (Spearman: R = -0.277) as well as for complexity 7 (Pearson: R = -0.281) were less pronounced and no trend was found for complexity 6. These findings are illustrated in figure 3.3.3., which shows that participants with a lower detour in the TSP task tended to do more gaze shifts during the Hanoi task.

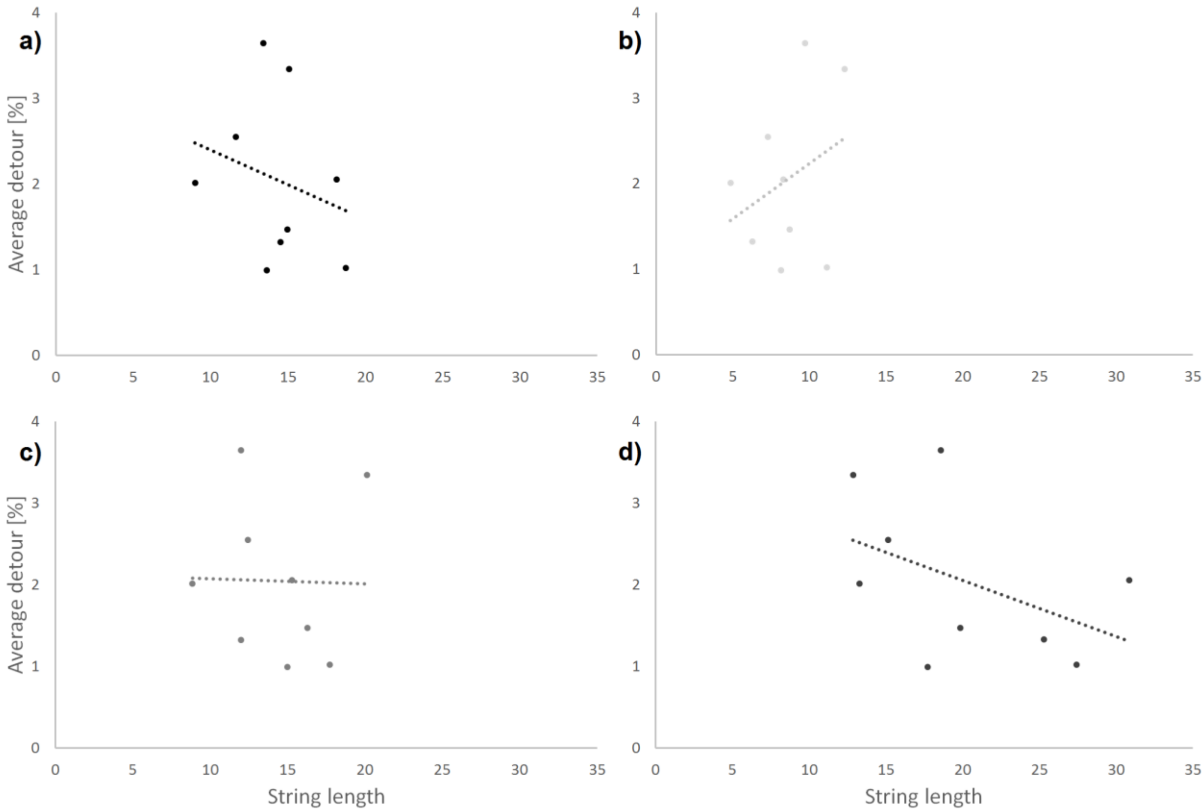


Figure 3.3.3.: Correlations between the average detours of the travelling salesperson task and the working string lengths of the Tower of Hanoi task. Similar to the correlations with the match rate, the strongest correlation could be found for the 4-move (b) problems (Spearman: R = -0.346). While there was no tendency for the 6-move (c) data, weak effects could be found for the 7-move (d) correlation (Pearson: R = -0.281) as well as the overall (a) data (Spearman: -0.277).

Finally, the only significant correlation between the average detours of the TSP and the error rates of the TOH was found for complexity 4 (Pearson: R = 0.574; $p < 0.05$; Spearman: R = 0.485; p = 0.05). Like the finding for error rates and the 2-back D-primes, this correlation needs to be looked at with some reservations, because there were only two errors made during complexity 4 of the TOH.

3.3.2 N-Back and TSP

Since the n-back task and TSP task were supposed to be assessments for different parameters, no significant correlation was expected between the data of these two tasks. While this assumption was true, a negative trend could be found for the 2-back data (Pearson: R = -0.425). Therefore, participants with high D-primes in the 2-back task showed a tendency to find shorter routes during the TSP.

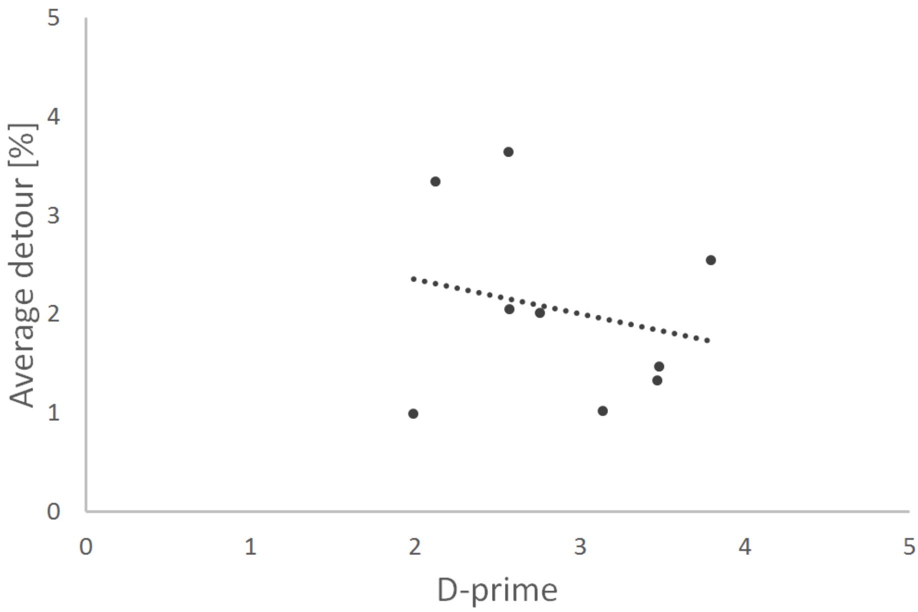


Figure 3.3.4.: Correlations between the average detours of the travelling salesperson task and the D-primes of the 2-back task. Contrary to previous expectations, a trend (Pearson: R = -0.425) of participants with better TSP results having higher D-primes during the 2-back task could be found. In contrast the 3-back data didn't show an effect.

4. Discussion

In conclusion, this study was conducted to investigate the role of eye movements during the planning process of a visual problem solving task. For this purpose, the Hanoi task used in this study, during which the participants' eye movements were recorded, was designed. After reducing those eye movements to fixations and later on gaze shifts between three working areas, the resulting path was compared to a predefined strategy. The results of the present study show that, while being very simplistic, the postulated strategy was used in many cases, which, especially for trials of complexity 7, showed a lower error rate.

Previous studies with similar experimental designs, like the Tower of London task, demonstrated that there are different phases of eye movements during visual problem solving. Hodgson et al. [2000], as well as Kaller et al. [2009] postulated the existence of different, sequentially phases, namely an information gathering or scouting phase and a solving phase. The current study shows that during the planning stage strategies can be used to enhance performance. For that an exemplary strategy was suggested, which was used in approximately one third of the trials across all participants, mainly for higher complexity problems. In addition, high variances were found across participants in the number of applications of the suggested strategy, the number of gaze shifts as well as the error rate. These findings give a hint on the existence of numerous different strategies which were most likely used by different participants. After establishing that a very basic eye movement strategy is used, further studies are necessary to look at the spectrum of possible other strategies.

The Hanoi task proved to be an adequate visual problem where distinct complexity levels can easily be defined. As expected the number of wrong answers increased with the number of moves necessary to solve the problem; this was also true for the total number of gaze shifts. Looking only at participants that used the postulated strategy it was found that an increase in complexity results in an increased number of fixations before the first match starts. In addition, for complexity 7 an increased number of trials with multiple matches was found. This implicates that strategy use is dependent on the difficulty of the problem and also that more complex problems lead to longer scouting phases as well as more repetitions to make sure the answer is correct. However, Kaller et al. [2010] cumulated evidence from a series of experiments and argued that an approximation of problem difficulty in terms of the minimum number of moves for goal attainment is too coarse for the underlying cognitive operations. He explained that to characterize problem difficulty more specifically a set of structural task parameters like the nature and number of solution paths, the patterns of intermediate and goal moves as well as the required search depth are required. Therefore, while the three different complexity levels often showed significantly different results, they might not represent three distinct difficulties. Further research is necessary to characterize different task levels which represent difficulty more precisely than the complexities used in this experiment.

Although, supplementary experiments failed to show significant effects, a number of strong tendencies could be found between TSP results and Hanoi data. These findings showed a tendency for participants with good TSP performance to also make less errors during the Hanoi task. In addition, participants with good TSP performance also tended to make more gaze shifts and had higher match rates during the Hanoi task. While none of those trends was significant, the TSP might still be a viable supplementary experiment for experiments like this. Further experiments with a higher number of participants might be able to show significant effects between TSP results and eye movement strategies.

Looking at correlations between the Hanoi data and the n-backs' D-primes, the only tendencies were in regards of the working string lengths. Participants with higher D-primes during the 2- and 3-back showed a tendency to make more gaze shifts during complexity 4 and 7 of the Hanoi task. This finding is unexpected since better n-back performance (Owen et al. [2005]) is an indicator for better working memory capabilities, which due to the trade-off between working memory and gaze shifts (Inamdar and Pomplun [2003]) is expected to reduce the need for gaze shifts. A possible explanation for this finding could be of a motivational nature. Highly motivated participants, that tend to do better during n-back tasks, might also spent more time to make sure they have found the correct answer during the Hanoi experiment, what would lead to an increase of gaze shifts and therefore working string length.

4.1. Conclusion

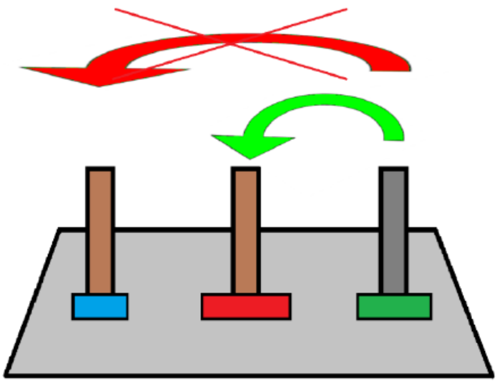
Overall this study succeeded in showing, on the basis of an example strategy, the strategic use of gaze shifts during the solving phase of visual problems as well as their influence on performance. Further studies are necessary to look into and potentially differentiate different strategies and evaluate their influence on performance. The choice of strategy as well as its performance might depend on the complexity of the problem, because strategy use was significantly lower in more simple tasks. Therefore, looking into a higher number of different complexities might yield interesting results. Lastly, the travelling salesperson task showed potential as a viable supplementary experiment, although a higher number of participants might be necessary to establish stronger effects.

A. Appendix

Towers of Hanoi

Aufgabe

- Ihre Aufgabe ist das Lösen von Rätseln, angelegt an das Spiel „Türme von Hanoi".
- Hierbei werden bei jedem Durchlauf mehrere, auf 3 Stäbe gestapelte, Scheiben angezeigt, die **nach bestimmten Regeln** auf einen Zielstab gebracht werden müssen.
- Die Scheiben sind verschiedenfarbig und haben unterschiedliche Durchmesser, von Groß nach Klein: **rot** → **grün** → **blau**
- Der **Zielstab** ist immer der ganz rechte Stab und **ist** grau gefärbt (die anderen **braun**).
- Ziel ist es die Scheiben nach Größe geordnet (unten Groß) auf den Zielstab zu stapeln
- Regeln zum Vesetzen der Scheiben
  1) Es darf pro Arbeitsschritt immer nur eine und nur die oberste Scheibe eines Stabes versetzt werden.
  2) Die zu versetzende Scheibe darf niemals auf einer kleineren Scheibe platziert werden (d.h. Anordnung von Groß nach Klein pro Stab).
- Das Ziel ist es, **schnellstmöglich die geringste Anzahl der zur Lösung benötigten Arbeitsschritte** zu bestimmen.



Ablauf

- Nach einer kurzen Kalibrierung des Eye-Trackers absolvieren Sie insgesamt 21 Durchläufe (Rätsel).
- Zu Beginn eines Durchlaufes wird in der Mitte des Bildschirms ein Kreuz angezeigt, dass für wenige Sekunden fixiert werden muss, bis das Rätsel eingeblendet wird.
- Dieses gilt es nun gedanklich zu lösen indem die minimal nötige Anzahl an Schritten bestimmt wird.
- Drücken Sie auf die Leertaste sobald Sie glauben die richtige Antwort ermittelt zu haben. Das Rätsel wird nun ausgeblendet und Sie sollen dem Versuchsleiter ihre Lösung verbal mitteilen.
- Durch Betätigen der Leertaste wird erneut ein Kreuz in der Mitte des Bildschirmes angezeigt und es startet der nächste Durchlauf.

Bitte versuchen Sie beim Lösen der Rätsel die minimale Anzahl an Lösungsschritten so genau und zügig wie möglich bestimmen!

Viel Spaß!

Figure A.1.: Written instructions for the Tower of Hanoi task.

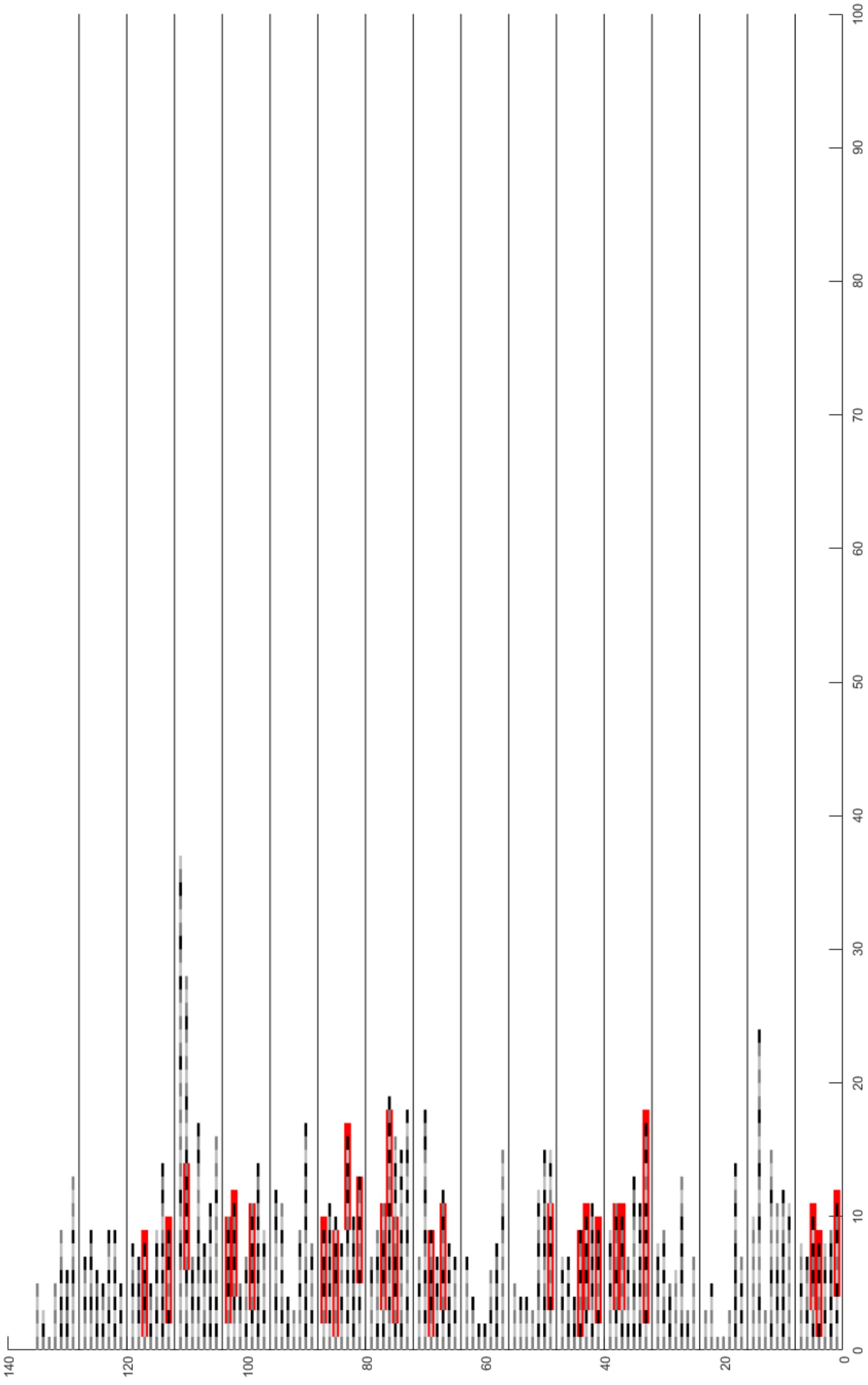


Figure A.2.: Working strings with matches of the 4-move Tower of Hanoi. The three zones are color coded in grey, with bright to dark resembling the zones one to three; matches are highlighted in red.

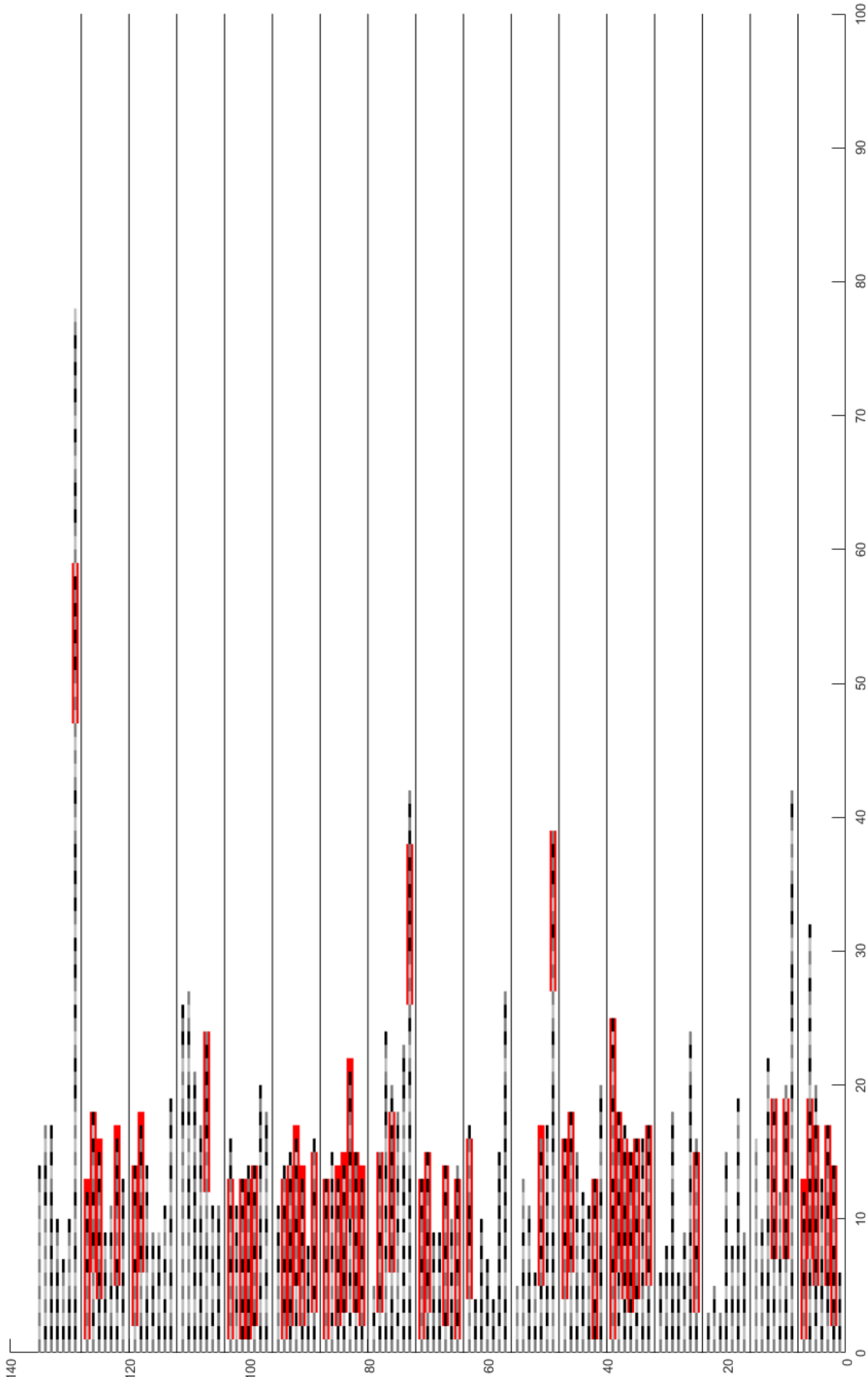


Figure A.3.: Working strings with matches of the 6-move Tower of Hanoi. The three zones are color coded in grey, with bright to dark resembling the zones one to three; matches are highlighted in red.

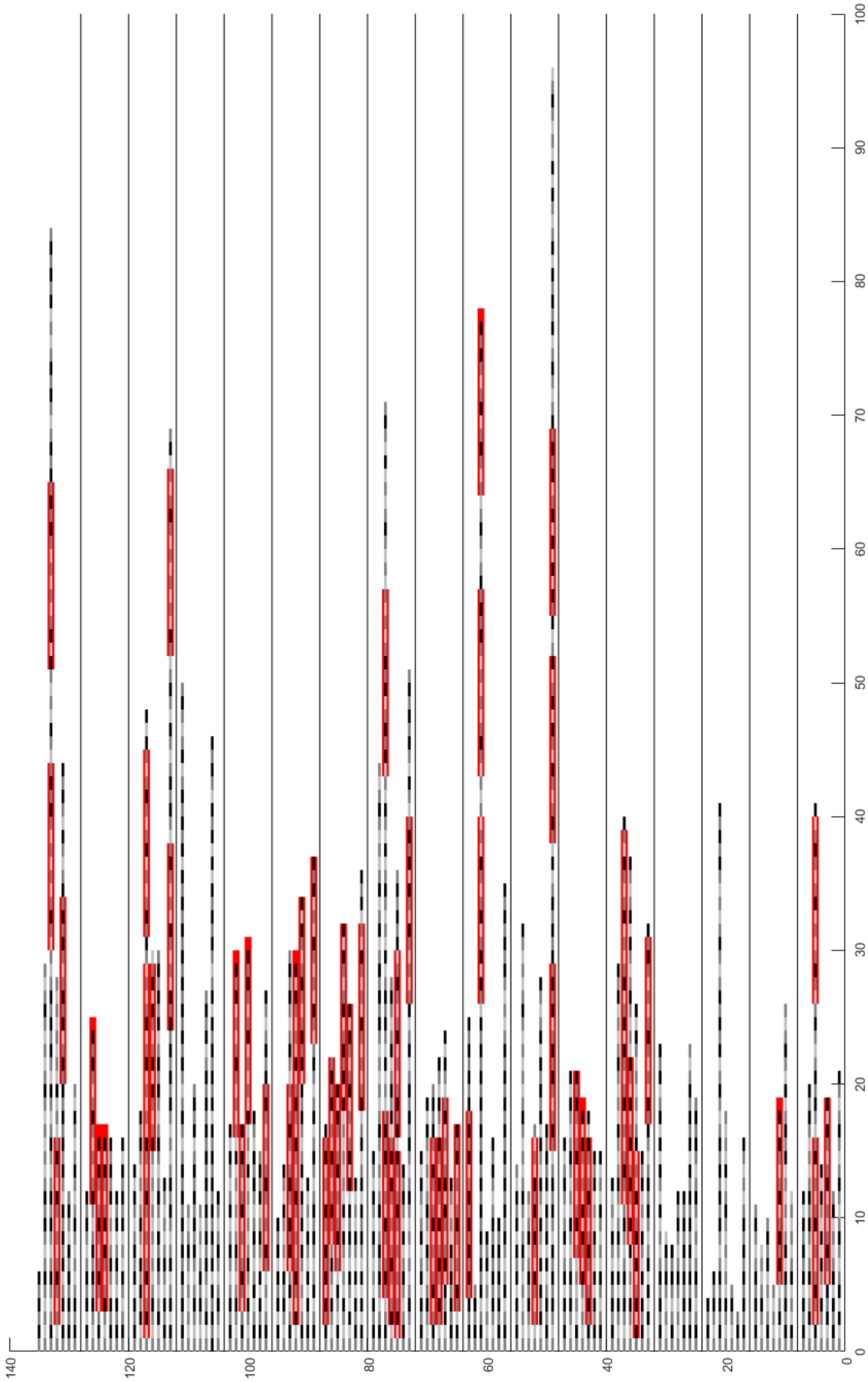


Figure A.4.: Working strings with matches of the 7-move Tower of Hanoi. The three zones are color coded in grey, with bright to dark resembling the zones one to three; matches are highlighted in red.

A.1. Statistical data

Table A.1.1.: Contingency table and Pearson chi-squared analysis between matches and errors of the TOH. Table a) shows the overall data, table b) the 4-move data, table c) the 6-move and table d) the 7-move data.

a)

115	20
171	51
$chi^2 = 3.507$ $p = 0.061$	

b)

29	1
88	1
$chi^2 = 0.663$ $p = 0.416$	

c)

41	10
54	14
$chi^2 = 0.017$ $p = 0.896$	

d)

44	12
32	31
$chi^2 = 9.913$ $p < 0.01$	

Table A.1.2.: Pearson and Spearman correlation data of the Tower of Hanoi match rate, error rate and string length.

		Error rate		String length	
		Pearson	Spearman	Pearson	Spearman
Match rate	4-move	$R = -0.114$ $p = 0.662$	$R = -0.075$ $p = 0.776$	$R = 0445$ $p = 0.073$	$R = 0.557$ $p < 0.05$
	6-move	$R = -0.460$ $p = 0.063$	$R = -0.446$ $p = 0.073$	$R = -0.024$ $p = 0.928$	$R = 0.037$ $p = 0.889$
	7-move	$R = -0.552$ $p < 0.05$	$R = -0.460$ $p = 0.063$	$R = 0.324$ $p = 0.204$	$R = 0.279$ $p = 0.279$
	Overall	$R = -0.568$ $p < 0.05$	$R = -0.477$ $p = 0.053$	$R = 0.306$ $p = 0.232$	$R = 0.350$ $p = 0.168$
String length	4-move	$R = -0,114$ $p = 0,662$	$R = -0,075$ $p = 0.776$		
	6-move	$R = 0.457$ $p = 0.066$	$R = 0.398$ $p = 0.114$		
	7-move	$R = -0.274$ $p = 0.288$	$R = -0.164$ $p = 0.529$		
	Overall	$R = -0.098$ $p = 0.708$	$R = -0.084$ $p = 0.748$		

Table A.1.3.: Pearson and Spearman correlation data of the supplementary experiments: n-back task and travelling salesperson task.

	3-back		TSP	
	Pearson	Spearman	Pearson	Spearman
2-back	R = 0.732 p < 0.001	R = 0.543 p < 0.05	R = -0.425 p = 0.089	R = -0.307 p = 0.231
TSP	R = -0.136 p = 0.604	R = -0.033 p = 0.900		

Table A.1.4.: Pearson and Spearman correlation data of the Tower of Hanoi parameters match rate, error rate and string length with the supplementary experiment data.

		2-back		3-back		TSP	
		Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
Error rate	4-move	R = -0.515 p < 0.05	R = -0.485 p < 0.05	R = -0.249 p = 0.336	R = -0.149 p = 0.568	R = 0.574 p < 0.05	R = 0.484 p < 0.05
	6-move	R = -0.122 p = 0.641	R = -0.094 p = 0.720	R = -0.256 p = 0.322	R = -0.142 p = 0.586	R = 0.355 p = 0.163	R = 0.274 p = 0.287
	7-move	R = 0.201 p = 0.440	R = 0.161 p = 0.538	R = 0.139 p = 0.595	R = 0.166 p = 0.525	R = 0.279 p = 0.278	R = 0.234 p = 0.366
	Overall	R = 0.041 p = 0.875	R = -0.044 p = 0.868	R = -0.024 p = 0.927	R = 0.019 p = 0.943	R = 0.403 p = 0.109	R = 0.359 p = 0.158
Match rate	4-move	R = 0.011 p = 0.968	R = -0.060 p = 0.819	R = 0.152 p = 0.561	R =0.093 p = 0.722	R = -0.464 p = 0.061	R = -0.504 p < 0.05
	6-move	R = -0.128 p = 0.624	R = -0,103 p = 0.694	R = 0.110 p = 0.676	R = 0.014 p = 0.956	R = -0.233 p = 0.369	R = -0.267 p = 0.301
	7-move	R = -0.163 p = 0.533	R = -0.238 p = 0.358	R = 0.015 p = 0.955	R = -0.004 p = 0.989	R = -0.126 p = 0.631	R = -0.189 p = 0.467
	Overall	R = -0.111 p = 0.671	R = -0.153 p = 0.559	R = 0.061 p = 0.816	R = -0.021 p = 0.937	R = -0.329 p = 0.197	R =-0.361 p = 0.154
String length	4-move	R = 0.252 p = 0.252	R = 0.158 p = 0.544	R = 0.197 p = 0.450	R = 0.152 p = 0.560	R = -0.307 p = 0.231	R = -0.346 p = 0.174
	6-move	R = 0.081 p = 0.757	R = 0.022 p = 0.935	R = 0.145 p = 0.578	R = 0.060 p = 0.820	R = 0.038 p = 0.884	R = -0.085 p = 0.747
	7-move	R = 0.438 p = 0.079	R = 0.437 p = 0.079	R = 0.405 p = 0.107	R = 0.382 p = 0.130	R = -0.281 p = 0.274	R = -0.255 p = 0.323
	Overall	R = 0.381 p = 0.132	R = 0.350 p = 0.169	R = 0.364 p = 0.151	R = 0.326 p = 0.201	R = -0.256 p = 0.321	R = -0.277 p = 0.282

A.2. Bibliography

Alvarez GA and Cavanagh P. 2004. *The capacity of visual short-term memory is set both by visual information load and by number of objects*. Psychological Science. 15:106–111.

Baddeley A. 1992. *Working memory*. Science. 255:556–559.

Baddeley A. 2003. *Working memory: Looking back and looking forward*. Nature Reviews Neuroscience. 4:829–839.

Baker SC, Rogers RD, Owen AM, Frith CD, Dolan RJ, Frackowiak RS and Robbins TW. 1996. *Neural systems engaged by planning: a PET study of the Tower of London task*. Neuropsychologia. 34:515–526.

Best BJ. 2006. *Cognitive Approaches to the Traveling Salesperson Problem: Perceptual Complexity that Produces Computational Simplicity*. AAAI Spring Symposium: Between a Rock and a Hard Place: Cognitive Science Principles Meet AI-Hard Problems. 11-16.

Carlin D, Bonerba J, Phipps M, Alexander G, Shapiro M and Grafman J. 2000. *Planning impairments in frontal lobe dementia and frontal lobe lesions*. Neuropsychologia. 38:655–665.

Cockburn J. 1995. *Performance on the Tower of London test after severe head injury*. Journal of the International Neuropsychological Society. 1:537–544.

Dagher A, Owen AM, Boecker H and Brooks DJ. 1999. *Mapping the network for planning: a correlation PET activation study with the TOL task*. Brain. 122:1973–1987.

Gilhooly KJ, Wynn V, Phillips LH, Logie RH and Della Sala S. 2002. Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual approach Thinking and Reasoning. 8:165–178.

Goel V. 2002. *Planning: Neural and Psychological*. Encyclopedia of Cognitive Science. London: Nature Publishing Group. 3:697-703.

Hodgson TL, Bajwa A, Owen AM and Kennard C. 2000. *The Strategic Control of Gaze Direction in the Tower of London Task*. Journal of Cognitive Neuroscience. 12(5):894-907.

Inamdar S and Pomplun M. 2003. *Comparative search reveals the tradeoff between eye movements and working memory use in visual tasks*. Proceedings of the Twenty-Fifth Annual Meeting of the Cognitive Science Society. 599–604.

Kaller CP, Unterrainer JM, Rahm B, Halsband U. 2004. *The impact of problem structure on planning: insights from the Tower of London task*. Cognitive Brain Research. 20:462–472.

Kaller CP, Rahm B, Bolkenius K and Unterrainer JM. 2009. *Eye movements and visuospatial problem solving: Identifying separable phases of complex cognition*. Psychophysiology. 46:818–830.

Kaller CP, Rahm B, Köstering L and Unterrainer JM. 2011. *Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks*. Behavioural Brain Research. 216:1-8.

Kaller CP, Rahm B, Spreer J, Weiller C and Unterrainer JM. 2011. *Dissociable Contributions of Left and Right Dorsolateral Prefrontal Cortex in Planning*. Cerebral Cortex. 21(2):307-317.

Kirchner WK. 1958. *Age differences in short-term retention of rapidly changing information*. Journal of Experimental Psychology. 55(4):352-358.

Lazeron RH, Rombouts SA, Machielsen WC, Scheltens P, Witter MP, Uylings HB and Barkhof F. 2000. *Visualizing brain activation during planning: the tower of London test adapted for functional MR imaging*. American Journal of Neuroradiology. 21:1407–1414.

Morris RG, Ahmed S, Syed GM and Toone BK. 1993. *Neural correlates of planning ability: frontal lobe activation during the Tower of London test*. Neuropsychologia. 31:1367–1378.

Morris RG, Rushe T, Woodruffe PW and Murray RM. 1995. *Problem solving in schizophrenia: a specific deficit in planning ability*. Schizophrenia Research. 14:235–246.

Nitschke K, Ruh N, Kappler S, Stahl C and Kaller CP. 2012. *Dissociable stages of problem solving (I): Temporal characteristics revealed by eye-movement analyses*. Brain and Cognition. 80(1):160-169.

Owen AM, Mcmillan KM, Laird AR and Bullmore E. 2005. *N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies*. Human Brain Mapping. 25(1):46-59.

Phillips W. 1974. *On the distinction between sensory storage and short term visual memory*. Perception & Psychophysics. 16:283–290.

Unterrainer J, Quiske K, Hoppe-Seyler K, Meier C, Müller C, Rahm B, Leonhart R and Halsband U. 2003. *Planning abilities and the Tower of London: Is this task measuring a discrete cognitive function?* Journal of Clinical and Experimental Neuropsychology. 26(6):846-856.

Ward G and Allport A. 1997. *Planning and problem-solving using the 5-disc Tower of London Task*. Quarterly Journal of Experimental Psychology. 50:49–78.

Ward G and Morris RG. 2005. *Introduction to the psychology of planning*. The Cognitive Psychology of Planning. Hove: Psychology Press. 1–34.

A.3. List of figures

A.4. List of tables