How Reasoning Ability, Working Memory Capacity and Conceptual Learning Interrelate: Behavioral and Neural Evidence

Author(s):
Rütsche, Bruno

Publication Date:
2017

Permanent Link:
https://doi.org/10.3929/ethz-b-000210709

Rights / License:
In Copyright - Non-Commercial Use Permitted
How Reasoning Ability, Working Memory Capacity and Conceptual Learning Interrelate: Behavioral and Neural Evidence

A thesis submitted to attain the degree of DOCTOR OF SCIENCES of ETH ZURICH (Dr. sc. ETH Zurich)

presented by
BRUNO RÜTSCHEN Master of Science UZH in Psychologie born on 20.06.1986 from Bütschwil-Ganterschwil, Bütschwil SG

accepted on the recommendation of
Prof. Dr. Elsbeth Stern
Prof. Dr. rer. nat Lutz Jäncke
Prof. Dr. Roland Grabner

2017
Acknowledgements

Writing this doctoral thesis has been quite a journey. It was both exciting, rewarding and the fulfillment of a dream, but at times - in all honesty - also a real test of patience, endurance, and self-discipline. Although only a single person is officially honored at the end, it has in fact been a group effort. It would never have been possible alone. Although there are many more persons who would deserve a thanks, there are a select few that I want to personally address.

First, I want to thank my supervisor Prof. Dr. Elsbeth Stern for making it possible to write my dissertation in her research group. She has widened my perspective with her knowledge throughout the years and provided me with guidance whenever I needed it. However, at the same time, she awarded me with an unparalleled freedom and trust to pursue my own interests. For that, I cannot be thankful enough.

All the people at the Chair for Research on Learning and Instruction also deserve a thanks for shaping the last few years. I am grateful for the many insightful and, at the same time, funny discussions; it would not have been the same without you.

Further thanks must go to both Prof. Dr. rer. nat Lutz Jäncke und Prof. Dr. Roland Grabner for being part of my doctoral committee and evaluating my thesis.

Although one can sometimes hardly believe it, there has been a time before being a doctoral student. Therefore, I want to thank my parents Elisabeth and Peter Rütsche for supporting me and allowing me to go my own way and brothers Marcel and Urs, well, for being brothers. Also thanks to all my friends (for a talk, sports, and/or "just" a beer); you know who you are.

Finally, although there really is no way to adequately put this into words, I want to thank my partner Silja Sollberger for her unwavering support, both professionally and personally. She has helped me countless times to shape my thoughts with her ingenuity and (nearly) always found a way to clear the confusion that at times seemed to overtake. But most importantly, thank you for all the wonderful experiences together, for putting everything into perspective and "simply" for being there.
Abstract

The goal of the present thesis was to improve our understanding of the interrelation between reasoning ability, working memory capacity, and conceptual learning. Three studies, investigating issues and applying methods from multiple research areas such as educational psychology, cognitive psychology, and cognitive neuroscience, were conducted in this regard.

The first empirical study was concerned with finding out why working memory capacity is so strongly related to intelligence, particularly reasoning ability. Depending on the theoretical position, interindividual differences in attention control, storage capacity or the ability for controlled retrieval from long-term memory are discussed as the driving mechanisms underlying this association. We wanted to add to this debate by comparing these theories not only from a behavioral but also from neural perspective. Therefore, we measured brain activity with electroencephalography (EEG) during working memory tasks that are representative for these theoretical accounts and related it to interindividual differences in reasoning ability. We found that neural activity indicative of all these mechanisms was related to reasoning ability, while behavioral performance was uninformative in this regard. We interpreted this in line with recent findings suggesting that the neurocognitive basis of working memory capacity and reasoning ability involves multiple mechanisms rather than a single one.

The second study used the same sample as the first study but focused on the role of these working memory mechanisms for conceptual learning in an educational context. Since the sample was recruited from a larger classroom study investigating physics learning in Swiss higher secondary schools, we could draw on an ecologically valid measure of conceptual learning. The statistical analysis showed that the performance in the task requiring controlled retrieval from long-term memory, after controlling for the processes involved in the other tasks, was uniquely predictive of conceptual learning. This confirms the crucial role of prior knowledge for learning, since the ability to specifically retrieve the relevant and disregard the irrelevant information might have
helped the development of a coherent conceptual knowledge base during the instruction. However, in contrast to the first study, there was no added value of neural measures over and above behavioral measures to - in this case - predict educational outcomes.

The main interest of the third study lay on the interplay of cognitive abilities and prior knowledge during conceptual learning. Although it is well-known in educational research that prior knowledge is the main factor influencing conceptual learning, studying its effect has played a secondary role in the category learning literature in cognitive psychology. This in turn severely restricts the applicability of these findings to real-world learning situations such as classrooms. In contrast, the complexity of learning in ecologically valid settings, such as the fact that cognitive ability and prior knowledge interact during development and are thus inevitably correlated, makes it difficult to disentangle specific factors. Thus, to bring educational and cognitive research on category and concept learning closer together, we took advantage of the control awarded by the experimental approach and developed a novel laboratory-based learning paradigm consisting of two subsequent learning phases: the first phase served to induce prior knowledge, which then had to be refined in the second phase. This allowed us to pinpoint the unique contributions of cognitive ability and prior knowledge during conceptual learning. We found that reasoning ability predicted learning in the first phase, while its effect in the second phase occurred indirectly over the effect on the first learning phase, i.e. over the acquired prior knowledge. This is an experimental corroboration of previous correlational findings showing that the impact of cognitive abilities on learning decreases once they have been invested in prior knowledge. With respect to learning in the second phase, prior knowledge was the strongest predictor, but working memory provided an additional effect. Furthermore, we also assessed learning strategies and found that participants who followed a rule abstraction rather than a rote learning approach achieved overall better learning gains.

Taken together, by drawing on issues and methods from multiple research areas, the three empirical studies provided novel insights about when and why reasoning ability, working memory capacity, and conceptual learning are related. The studies, however, have not only improved our understanding of these associations, but have also yielded new research questions and possibilities. Accordingly, it is my hope that the present work serves as an inspiration and starting point for future research.
Zusammenfassung


Die erste empirische Studie befasste sich mit der Frage, wieso Arbeitsgedächtniskapazität und schlussfolgerndes Denken so stark zusammenhängen. Für diese enge Beziehung werden je nach theoretischem Hintergrund interindividuelle Unterschiede in der Aufmerksamkeitssteuerung, der Speicherkapazität oder der Fähigkeit zum kontrollierten Abruf aus dem Langzeitgedächtnis verantwortlich gemacht. Wir wollten zu dieser Debatte beitragen, indem wir diese Theorien nicht nur auf der Verhaltensebene, sondern auch auf neuronaler Ebene miteinander verglichen. Aus diesem Grund erfassten wir mit einem Elektroenzephalogramm (EEG) die Hirnaktivität während kognitiven Aufgaben, welche die oben genannten kognitiven Mechanismen besonders erforderten. Es zeigte sich, dass die Hirnaktivität während all dieser Mechanismen mit der individuellen Fähigkeit zu schlussfolgerndem Denken zusammenhing, während die Leistung in den Aufgaben diesbezüglich nicht informativ war. Diesen Befund interpretierten wir in Übereinstimmung mit neueren Erkenntnissen so, dass der Zusammenhang zwischen Arbeitsgedächtniskapazität und schlussfolgerndem Denken aufgrund mehrerer Mechanismen und nicht eines einzelnen Mechanismus zu Stande kommt.

Die zweite Studie verwendete die gleiche Stichprobe wie die erste Studie, untersuchte jedoch den Einfluss dieser Arbeitsgedächtnismechanismen auf konzeptuelles Lernen in einem schulischen Kontext. Wir konnten auf ein ökologisch valides Mass für konzeptuelles Lernen zurückgreifen, weil die Teilnehmer zuvor an einer größeren Klassenraum-Studie teilgenommen hatten, welche zum Ziel hatte, Physiklernen in Schweizer Gymnasien zu untersuchen. Die statistische Analyse zeigte, dass die Leistung in der Aufgabe zum kontrollierten Abruf aus dem Langzeitgedächtnis, nachdem für die Prozesse der an-
Zusammenfassung
deren Aufgaben kontrolliert wurde, besonders prädiktiv für konzeptuelles Lernen war. Die Fähigkeit, sich auf die relevanten Informationen zu beschränken und irrelevante Inhalte zu ignorieren, schien somit den Aufbau einer kohärenten Wissensbasis während dem Unterricht zu begünstigen. Dieser Befund unterstreicht die zentrale Rolle, welche die Aktivierung von Vorwissen für erfolgreiches Lernen spielt. Im Gegensatz zur ersten Studie konnte jedoch kein zusätzlicher, über die Verhaltensaufflage hinausgehender Nutzen der Berücksichtigung neuronaler Masse für die Vorhersage des Lernerfolgs festgestellt werden.

Zusammenfassung

zu abstrahieren, generell bessere Leistungen erzielten als solche, die eher auswendig gelernt hatten.

Insgesamt konnten in den drei durchgeführten Studien durch die Anwendung von Methoden mehrerer Forschungsgebiete neuartige Erkenntnisse darüber gewonnen werden, wann und wieso schlussfolgerndes Denken, Arbeitsgedächtniskapazität und konzeptuelles Lernen zusammenhängen. Die Studien haben jedoch nicht nur das Verständnis dieser Beziehungen verbessert, sondern auch neue Forschungsfragen und -möglichkeiten hervorgebracht. Der Autor hofft deshalb, dass die vorliegende Arbeit als Inspiration und Ausgangspunkt für weitere Forschung dient.
Contents

3.1.3 The Current Study ........................................... 40
3.2 Method .......................................................... 42
  3.2.1 Sample ....................................................... 42
  3.2.2 Procedure .................................................... 43
  3.2.3 Measures ...................................................... 43
  3.2.4 EEG Recording and Preprocessing ....................... 49
  3.2.5 Statistical Analyses ....................................... 51
3.3 Results .......................................................... 52
  3.3.1 Behavioral Data ............................................ 52
  3.3.2 Neural Data ................................................ 54
3.4 Discussion ..................................................... 59
3.5 Acknowledgments .............................................. 64
References .......................................................... 64

4 The Relation Between Working Memory and Conceptual Learning: An EEG Study ........................................... 74

  Abstract .......................................................... 74
  4.1 Introduction ................................................... 75
    4.1.1 The Role of Working Memory for Learning .......... 75
    4.1.2 The Added Value of Neuroimaging .................... 79
    4.1.3 The Present Study ....................................... 81
  4.2 Method ........................................................ 82
    4.2.1 Participants ............................................... 82
    4.2.2 Procedure .................................................. 83
    4.2.3 Measures ................................................... 83
    4.2.4 Electroencephalography (EEG) ......................... 86
    4.2.5 Statistical Analysis ..................................... 87
  4.3 Results ........................................................ 89
    4.3.1 Performance ............................................... 89
    4.3.2 EEG Power / Connectivity ............................... 91
  4.4 Discussion ................................................... 93
  4.5 Acknowledgments ............................................. 96
References .......................................................... 96
## Contents

5 Constructing and Refining Relational Categories: Unraveling the Interplay between Prior Knowledge, Reasoning Ability, Working Memory Capacity, and Strategy 106

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>106</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>107</td>
</tr>
<tr>
<td>5.1.1 Category Learning and Conceptual Change</td>
<td>107</td>
</tr>
<tr>
<td>5.1.2 Interindividual Differences in Relational Category Learning</td>
<td>108</td>
</tr>
<tr>
<td>5.1.3 The Current Study</td>
<td>109</td>
</tr>
<tr>
<td>5.2 Method</td>
<td>111</td>
</tr>
<tr>
<td>5.2.1 Participants</td>
<td>111</td>
</tr>
<tr>
<td>5.2.2 Procedure</td>
<td>112</td>
</tr>
<tr>
<td>5.2.3 Measures</td>
<td>112</td>
</tr>
<tr>
<td>5.3 Results</td>
<td>117</td>
</tr>
<tr>
<td>5.3.1 Learning Phases</td>
<td>117</td>
</tr>
<tr>
<td>5.3.2 Assessments 1 and 2</td>
<td>118</td>
</tr>
<tr>
<td>5.4 Discussion</td>
<td>126</td>
</tr>
<tr>
<td>5.5 Acknowledgments</td>
<td>128</td>
</tr>
<tr>
<td>Appendix</td>
<td>129</td>
</tr>
<tr>
<td>References</td>
<td>134</td>
</tr>
</tbody>
</table>

6 General Discussion 138

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Integrative Summary of the Main Findings</td>
<td>138</td>
</tr>
<tr>
<td>6.2 Strengths and Limitations</td>
<td>141</td>
</tr>
<tr>
<td>6.3 Outlook and Implications</td>
<td>142</td>
</tr>
<tr>
<td>6.4 Conclusion</td>
<td>144</td>
</tr>
<tr>
<td>References</td>
<td>145</td>
</tr>
</tbody>
</table>

Appendix 149

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Sheet: Study 1 / 2</td>
<td>150</td>
</tr>
<tr>
<td>Consent Form: Study 1 / 2</td>
<td>152</td>
</tr>
<tr>
<td>Information Sheet: Study 3</td>
<td>153</td>
</tr>
<tr>
<td>Consent Form: Study 3</td>
<td>155</td>
</tr>
<tr>
<td>Task Instructions: Stroop</td>
<td>156</td>
</tr>
</tbody>
</table>
Contents

Task Instructions: Visual Arrays ............................................. 158
Task Instructions: Secondary Memory ................................. 160
Task Instructions: Updating ............................................... 163
Task Instructions: Category Learning - Phase 1 .................... 166
Task Instructions: Category Learning - Phase 2 .................... 170
List of Figures

Figure 2.1 Example item of a matrices test in the style of Raven’s Advanced Progressive Matrices (RAPM) .................................................. 9
Figure 2.2 Multi-component model of working memory .............................. 12
Figure 2.3 Embedded-processes model of working memory ........................ 15
Figure 3.1 Trial structure of the Stroop task. ........................................ 45
Figure 3.2 Trial structure of the visual arrays task. ................................. 46
Figure 3.3 Trial structure of the secondary memory task. .......................... 48
Figure 3.4 Trial structure of the updating task. ..................................... 49
Figure 3.5 Schematic of the EEG layout. ............................................. 50
Figure 3.6 EEG results of the Stroop task ............................................ 55
Figure 3.7 EEG results of the visual arrays task ................................... 56
Figure 3.8 EEG results of the secondary memory task ............................ 57
Figure 3.9 EEG results of the updating task ....................................... 58
Figure 4.1 Sample item of the basic Mechanics Conceptual Understanding (bMCU) 84
Figure 4.2 Schematic depiction of the working memory tasks. ................. 86
Figure 5.1 Illustration of the relational category system ........................... 114
Figure 5.2 Sample items for the learning phases and assessments. ............. 115
Figure 5.3 Solution rates in both learning phases. ................................. 117
Figure 5.4 Results of the structural equation models estimated for each type of transfer item. .................................................. 119
List of Tables

Table 3.1 Descriptive statistics of the behavioral data. ........................... 53
Table 3.2 Correlation matrix of the behavioral data. ............................... 54
Table 4.1 Descriptive statistics of the behavioral measures. ....................... 89
Table 4.2 Correlation matrix of all behavioral measures. .......................... 90
Table 4.3 Results of the regression model predicting bMCU post by working memory task performance. .................................................. 91
Table 4.4 Parameter estimates of the relation between bMCU post and functional network parameters for the theta (3-6 Hz) and alpha bands (8-12 Hz). 92
Table 5.1 Descriptive statistics of all observed measures and their correlations. 120
Table 5.2 Direct effects in the structural equation models for each type of transfer item. ............................................................. 123
Table 5.3 Indirect effects in the structural equation models conducted for each type of transfer item. ....................................................... 124
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian Information Criterion</td>
</tr>
<tr>
<td>bMCU</td>
<td>basic Mechanics Conceptual Understanding</td>
</tr>
<tr>
<td>CHC</td>
<td>Carroll-Horn-Cattell</td>
</tr>
<tr>
<td>Cw</td>
<td>Mean clustering coefficient</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>FIML</td>
<td>Full Information Maximum Likelihood</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>ICA</td>
<td>Independent Component Analysis</td>
</tr>
<tr>
<td>Lw</td>
<td>Characteristic path length</td>
</tr>
<tr>
<td>M</td>
<td>Mean</td>
</tr>
<tr>
<td>P-FIT</td>
<td>Parieto-Frontal Integration Theory</td>
</tr>
<tr>
<td>RAPM</td>
<td>Raven’s Advanced Progressive Matrices</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SE</td>
<td>Standard Error</td>
</tr>
<tr>
<td>SSR</td>
<td>Strategy Self-Reports</td>
</tr>
<tr>
<td>STEM</td>
<td>Science, Technology, Engineering, Mathematics</td>
</tr>
<tr>
<td>WMC</td>
<td>Working Memory Capacity</td>
</tr>
<tr>
<td>wPLI</td>
<td>weighted Phase Lag Index</td>
</tr>
</tbody>
</table>
1 Overview

People differ. They differ in many respects, including their knowledge, their abilities, their personalities, and their interests. Although this observation is rather trivial, understanding these differences is far from that; it really lies at the core of understanding the human mind. This thesis presents three studies with the overall goal to shed more light on what makes humans different, with a focus on the interrelationship between reasoning ability, working memory capacity (WMC), and conceptual learning.

Psychometric intelligence is one of the constructs with the longest and most successful history in psychological research to capture such interindividual differences. Intelligence can be defined in very general terms as the "ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought" (Neisser et al., 1996). Despite some ongoing debates over the structure of intelligence, fluid intelligence or reasoning ability is a - if not the - core component of contemporary theories of human intelligence. It generally refers to the ability to solve novel problems by reasoning and is assumed to be largely independent of previous experience and knowledge (Cattell, 1963).

Another psychological construct with respect to cognitive ability that has gained wide popularity in the recent decades is working memory. Working memory has been developed in experimental and cognitive psychology and generally describes a cognitive system that serves to maintain information in an accessible state for goal-directed processing (Baddeley, 2012; Baddeley & Hitch, 1974; Cowan, 1999). Naturally following from this characterization, working memory is ubiquitous and plays a crucial role in many areas of life. It is involved in simple everyday tasks such as memorizing a shopping list, but is also strongly taxed in school when learning new vocabulary, performing mental calculations or acquiring knowledge about scientific concepts (e.g., buoyancy, Newtonian force, Darwin’s biological theory of evolution). Interindividual differences in WMC and reasoning ability have been shown to be closely related, which
Overview

is why studying this relationship is seen as a unique opportunity to gain a better understanding of intelligence (e.g. Kane, Hambrick, & Conway, 2005; Kyllonen & Christal, 1990; Oberauer, Schulze, Wilhelm, & Süß, 2005). However, the cognitive basis of this association is still a matter of debate. Depending on the theoretical assumptions made about the nature of WMC, the importance of executive functions and attentional control, storage capacity, or retrieval ability is emphasized (Cowan et al., 2005; Engle, 2002; Unsworth & Engle, 2007). The goal of the first study presented in this thesis was to compare these theoretical accounts.

Furthermore, according to these definitions, both reasoning ability and working memory are innately linked to learning. The association between cognitive abilities and learning is so intuitive that intelligence was even equated with the "ability to learn" in early definitions (e.g. Buckingham, 1921). Surprisingly, however, the history of research on this topic is less clear than one would expect. For example, early influential research by Woodrow (1938, 1946) found that ability scores were poor predictors of learning in several simple laboratory tasks, which led him to deny any meaningful relation between the two. This provoked much criticism in the following years and, as a consequence, several factors have been pointed out that affect the ability-learning relationship, including the complexity of the task (for an overview see Lohman, 1999). Indeed, the relationship has been shown to be stronger when the task not only depends on simple associative learning (as typical for learning tasks applied in the laboratory) but requires that one goes beyond the given information to abstract and integrate knowledge (Ackerman, 2011; Alexander & Smales, 1997; Primi, Ferrão, & Almeida, 2010; Voelkle, Wittmann, & Ackerman, 2006; Williams & Pearlberg, 2006). This is exactly the situation students face in today’s education, especially in the STEM (Science, Technology, Engineering, Mathematics) fields, where students have to gradually acquire a wealth of highly complex and interconnected scientific concepts in a few years of schooling (Resnick, 2010). Unfortunately, though, many students leave school without a deeper understanding of even basic science concepts such as Newton’s laws of motion (e.g., Carey, 2000; Halloun & Hestenes, 1985; Hestenes, Wells, & Swackhamer, 1992).

The second study of the present thesis thus focused on the role of working memory as a resource for successful conceptual learning in an educational context (i.e., classrooms). The same sample was used in the first and second study, but the focus was on different constructs (i.e., on conceptual learning rather than on reasoning ability). One distin-
guishing feature of both studies is that we applied a neuroscientific method, specifically electroencephalography (EEG), to measure brain activity during working memory performance. This additional level of information makes it possible to reveal processes that are otherwise invisible and thus provide unique insights.

A further factor known to strongly influence future learning is domain-specific knowledge or prior knowledge. There is a wealth of findings from educational psychology that prior knowledge is the most important driver of learning, typically outweighing the influence of cognitive abilities (e.g., Ackerman, 2007; Hambrick & Meinz, 2011; Schneider, Körkel, & Weinert, 1989; Stern, 1999, 2015; Tricot & Sweller, 2013; Weinert, Helmke, & Schneider, 1990). However, the specific influences of cognitive abilities and prior knowledge on learning are difficult to identify in real-world settings, because the (inevitably present) differences in prior knowledge are at least partially the consequence of the investment of cognitive abilities. In contrast, laboratory-based experimental studies on category or conceptual learning typically reduce the influence of prior knowledge by using highly abstract and artificial learning materials (Goldwater & Schalk, 2016). Although this diminishes the ecological validity of the findings, the experimental setting has the advantage of providing good control over the learning process and situation (in contrast to the complexity of factors influencing school learning). In the third study, we attempted to bridge the gap between educational and experimental studies on category and conceptual learning. To this aim, we developed a novel category learning paradigm consisting of two learning phases, i.e. participants constructed prior knowledge in a first learning phase and refined it in a second learning phase. This setup made it possible to identify the specific contributions of cognitive abilities and prior knowledge (and in this case also strategic approaches) during conceptual learning.

The rest of this thesis is structured as follows: First, the general introduction provides the relevant theoretical background and concludes with the research questions (chapter 2). Afterwards, the three empirical studies are described in three self-contained chapters (chapters 3, 4 and 5). Finally, a summary and an integrative discussion of the main findings as well as an overall conclusion are provided (chapter 6).
References


References


References


2 General Introduction

2.1 Intelligence and Reasoning Ability

An important step to better understand the construct intelligence is to be broadly acquainted with its history. At the end of the 19th century, Francis Galton ascribed individual differences in cognitive ability to basic processes such as finer sensory discrimination and faster reaction times (for a detailed account see Mackintosh, 2011). However, these assumptions have been discounted by early empirical findings; the different measures did not correlate well with each other nor were they related to school grades (e.g., Wissler, 1901). In contrast, Alfred Binet’s attempts to measure intelligence were more successful. He attributed intelligence to higher cognitive functions such as attention and memory and devised more complex tests to measure children’s ability to cope with everyday problems (e.g., counting or naming the months of the year). The tests were, in line with their main aim of the development, highly sensitive in distinguishing normally developing and mentally retarded children. Over the next decades, intelligence research was generally concerned with identifying the structure of intelligence, mostly by applying regression and factor-analytic approaches to characterize the relationships between cognitive tests in the tradition of Binet. Famously, Spearman (1927) argued that the generally positive correlation between the performance on most cognitive tests (the ”positive manifold”) is due to a common underlying factor (the general factor or $g$ factor). He assumed that $g$ reflects some unspecified mental energy that causes the intercorrelations between tests. In contrast, Thurstone (1938) initially denied the relevance of such a factor but instead advocated seven independent “primary mental abilities” such as reasoning, memory, numbers, and verbal comprehension.

Another prominent distinction, based on the work of Cattell (1963, 1971), is the one between crystallized and fluid intelligence. Crystallized intelligence refers to knowledge and skills acquired through education and cultural influences. It is typically assessed with questions drawing on common knowledge in various domains (such as lan-
2.1 Intelligence and Reasoning Ability

Fluid intelligence reflects reasoning ability, which is used to solve problems by manipulating and processing new information\(^1\). It is thought to be largely independent of previous knowledge and is typically measured with content-poor reasoning tasks (e.g., Matrix completion, verbal analogies, or number series). The matrices tests developed by John Raven are arguably the most used measures of fluid intelligence. Figure 2.1 shows a sample item that follows the style of Raven’s Advanced Progressive Matrices (RAPM; Raven, Raven, & Cout, 2003). The objective is to infer one or multiple rules underlying a 3×3 matrix of geometrical figures and complete the pattern by selecting the correct answer out of several potential candidates. The separation of intellectual ability into fluid and crystallized intelligence has proven to be particularly fruitful with respect to developmental and educational research. For example, both factors have been shown to have differential patterns across the lifespan: While fluid intelligence generally decreases with age, crystallized intelligence increases (Horn & Cattell, 1967). Furthermore, fluid and crystallized intelligence are assumed to be developmentally related. Specifically, according to Cattell’s investment theory of intelligence (e.g., Cattell, 1963, 1971), fluid intelligence is invested in the acquisition of knowledge and skills, or conversely, crystallized intelligence is the result of past reasoning activities (see also Schweizer & Koch, 2001; Valentin Kvist & Gustafsson, 2008).

At the end of the last century, Carroll (1993) reanalyzed a wealth of data on cognitive abilities and provided a consolidation of Spearman’s, Thurstone’s, and Cattell’s theories. He concluded that intellectual ability consists of three levels: At the top level the \(g\) factor reappears, capturing the common variance of several second level factors (such as fluid intelligence, crystallized intelligence, processing speed, short-term memory etc.), which are again subdivided into specific factors on the lowest level (Carroll, 1993). The model, sometimes referred to as the Carroll-Horn-Cattell (or CHC) model, is now seen as the state-of-the-art structural model of intelligence (McGrew, 2009).

From a practical perspective, the history of intelligence research has been a success: The construct demonstrates excellent statistical properties, is highly stable across the lifetime and among the best predictors of academic achievement and job performance (Deary, Strand, Smith, & Fernandes, 2007; Deary, Whiteman, Starr, Whalley, & Fox, 2007).

\(^1\)The terms fluid intelligence and reasoning ability are oftentimes used interchangeably. In the three studies described in the present thesis, we will consistently use the term reasoning ability.
2.1 Intelligence and Reasoning Ability

Figure 2.1: Example item of a matrices test in the style of Raven’s Advanced Progressive Matrices (RAPM; Raven, Raven, & Cout, 2003). Answer alternative 6 is correct.

2004; Neisser et al., 1996; Schmidt & Hunter, 1998). However, from a theoretical perspective, intelligence is still a relatively fuzzy concept that is difficult to define. This can partially be related to the fact that structural models of intelligence mainly try to capture statistical relationships between a large battery of different tasks that, however, have not been constructed or selected following a coherent theoretical model of intelligence (Oberauer, Schulze, Wilhelm, & Süß, 2005). For example, the CHC model neglects developmental aspects such as the causal influence of reasoning ability on the acquisition of knowledge, as suggested in Cattell’s investment theory. The situation is further complicated by the fact that fluid intelligence tests, which were specifically designed to be independent of prior experience (see Figure 2.1), are still affected by education level and culture (Rosselli & Ardila, 2003). These tests not only measure biologically determined fluid abilities as intended, but are in parts also influenced by acquired knowledge.
2.1 Intelligence and Reasoning Ability

Therefore, reasoning ability is at the same time an important predictor for academic achievement and one of the central outcomes of education (Snow, 1996). Furthermore, although the statistical concept of $g$ is generally accepted, the question remains whether there is a single psychological process underlying it. Bartholomew, Deary, and Lawn (2009) reviewed theories on human intelligence and came to the conclusion that, although a $g$ factor reflecting a common cognitive process is a sufficient explanation for the positive manifold, it is not necessary. Other ideas, such as that different tasks involve a great number of cognitive processes and that there is at least some overlap between these processes across tasks, provide equally viable explanations (see also van der Maas et al., 2006).

The lack of a theory of intelligence arguably inspired novel approaches at the end of the last century. These new research strands were more rooted in experimental and cognitive psychology following an information processing perspective, as compared with the psychometric view that had dominated intelligence research until then. The general idea was to apply a reductionist approach and decompose intelligence into its elementary cognitive components. In one account, reviving the early idea of Galton (see p. 7), intellectual abilities were argued to be the result of the speed and efficiency of information processing (for a review see Jensen, 2006). This is typically assessed in so-called elementary cognitive tasks which require simple perceptual and semantic discrimination between stimuli. For example, Jensen and Munro (1979) used a novel apparatus to measure reaction times: It consisted of several buttons that were all equidistant from a "home" position as well as lights over each button. After one of the lights was randomly turned on, the subjects had to press the corresponding button as quickly as possible. In contrast to the then commonly held notion, reaction times in this task were indeed related to intelligence (about $r = -.4$). Since then, a wealth of studies have confirmed the consistent but modest relationship between intelligence and processing speed (Sheppard & Vernon, 2008), but the exact nature of this association has not yet been resolved (Nettelbeck, 2011). Furthermore, since a large portion of variance has remained unexplained, the currently accepted view is that intelligence cannot be completely reduced to processing speed and other factors must be at play. One such factor discussed with respect to the cognitive basis of intelligence is working memory (e.g., Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009; Oberauer, Süß, Wilhelm, & Wittmann, 2008). This will be the topic of the next chapter.
2.2 Working Memory

Working memory is arguably among the most influential and important concepts of cognitive psychology, maybe even the whole of psychology. At the time of writing (June 2017), searching the PsycInfo database for papers containing "working memory" in either the title or the abstract returned a massive amount of 26’712 results, while in Google Scholar over 2’040’000 results were found. Historically, although general ideas about the necessity of being able to hold information in mind go back much further (for a short overview see Cowan, 2014), the term "working memory" first appeared in the seminal book of Miller, Galanter, and Pribram (1960). In a general attempt to shine a light on the black box of behaviorism, they recognized the need of a working memory as a space for the planning and execution of behavior.

The working memory model that shaped most of the research in this area, however, has been the multi-component model developed by Baddeley and Hitch (1974). One central assumption of the model is that short-term memory is separate from long-term memory, in line with the at that time dominant model of short-term memory by Atkinson and Shiffrin (1968). However, in contrast to that model, Baddeley and Hitch (1974) argued that the short-term store is not unitary but consists of multiple components. This argumentation was based on several findings (see Baddeley, 2012; Baddeley & Hitch, 1974; Baddeley, 1992). First, healthy participants were able to adequately perform complex cognitive tasks (e.g., reasoning) even when their short-term store was strongly taxed by a concurrent task (e.g., remembering digits). If there was only a single store available, the performance should have been completely disrupted. Second, in a similar vein, patients with short-term memory deficits were still able to perform many complex tasks on a normal level. Third, there was evidence for a double dissociation between processing and storage tasks based on the domain: Phonological processing interfered with phonological but not visual-spatial storage and visual-spatial processing interfered with visual-spatial storage but not phonological storage.

These results, among others, led Baddeley and Hitch (1974) to propose a multi-component system that consists of a control system, the "central executive", and two independent storage systems (see Figure 2.2). The "phonological loop" was argued to be responsible for the maintenance of verbal/phonological information via a subvocal rehearsal process, while the "visual-spatial sketchpad" was assumed to be involved in
the maintenance of visual information. Later, the "episodic buffer" was introduced to, among other issues, account for the fact that long-term memory can greatly influence working memory performance (Baddeley, 2000). The episodic buffer is thought of as a space where representations from different sources can be integrated, generally acting as a buffer between the other components as well as long-term memory. For example, the well-known finding that more words can be recalled when they form a meaningful sentences (as compared to a sequence of unrelated words) can be explained by the supposed function of the episodic buffer to integrate or "chunk" the words.

Figure 2.2: Multi-component model of working memory. Figure adapted from Baddeley (2000). LTM = Long-Term Memory.

Overall, this conceptualization of short-term memory moved away from a simple store to a system of components serving the simultaneous storage and processing of information. Therefore, tasks that involve both storage and processing have been developed to measure the construct, such as the complex span tasks in which the maintenance of items for later recall is interspersed with an additional (irrelevant) processing task. For example, the reading span task designed by Daneman and Carpenter (1980) requires that participants read aloud a series of sentences (processing), while maintaining the last word of each sentence for later recall (storage). In the operation span task, participants alternate between verifying simple arithmetic equations and remembering letters (Turner & Engle, 1989). In line with the assumption of the multi-component model, complex span tasks show stronger correlations with demanding cognitive tasks.
2.2 Working Memory

(such as reading and language comprehension or reasoning) than simple span tasks that arguably involve only storage (such as the digit or word span; see e.g., Daneman & Carpenter, 1980; Kane et al., 2004; Turner & Engle, 1989). However, while tasks involving simultaneous storage and processing are a good way to measure WMC, recently there has been an increasing realization that tasks involving either pure storage or processing are just as adequate, since they seem to draw largely on the same mechanisms (e.g., Conway, Getz, Macnamara, & Engel de Abreu, 2011; Oberauer et al., 2008).

Although the multi-component model has drastically advanced the understanding of working memory, several of its basic assumptions have been challenged, in particular by neuroscientific research. For example, the strict dissociation between short-term and long-term memory seems no longer as tenable, since both have been shown to involve similar neural substrates, including the hippocampus (i.e., the classical long-term memory; Jonides et al., 2008; Ranganath & Blumenfeld, 2005). Furthermore, there has been increasing evidence that the simple dichotomy of verbal versus visual-spatial buffers is insufﬁcient and that more domains ought to be distinguished (e.g., touch, olfaction; Postle, 2006). In contrast to the multi-component model, so-called state-based models of working memory can account for these ﬁndings in a more straightforward and parsimonious way. Generally, state-based models conceptualize working memory as an attentional system that directly acts upon long-term representations by increasing or decreasing their activation states (e.g., Cowan, 1988, 1999; Oberauer, 2002) Thus, in contrast to dual-store models in which short-term memory represents a gateway to long-term memory (Atkinson & Shiffrin, 1968; Baddeley, 2000), there is no structural division between working and long-term memory. Furthermore, instead of introducing additional boxes into the model to account for the many potential domain-speciﬁc buffers, state-based models assume that all information is part of a general set of activated long-term representations (Cowan, 2014). The general assumption of state-based models, i.e. that working memory is achieved by the allocation of attention to internal representations in long-term memory, is well in line with current neuroscientiﬁc data. That is, when maintaining information in working memory, there is increasing evidence that frontal areas, which are assumed to mainly subserve domain-general attentional processes, continuously interact with posterior areas, which are thought to store domain-speciﬁc representations (e.g., Jonides et al., 2008; Postle, 2006; Postle,

\[^2\text{However, Norris (2017) has recently noted potential problems with these findings.}\]
2.2 Working Memory


The most general and well-known state-based model is the embedded-processes model of Nelson Cowan (1988, 1999). According to this model (see Figure 2.3), working memory consists of a central executive responsible for directing attention and a short-term store consisting of activated long-term memory representations. And within that activated portion of long-term memory, a limited subset of representations with particularly high activation levels are assumed to be in the focus of attention. The main function of the focus of attention is to make information immediately accessible for active cognitive processing such as integrating or chunking elements into more complex structures (Halford, Cowan, & Andrews, 2007; Oberauer, 2009). In the embedded-processes model and in state-based models in general, attention is a crucial aspect of a functioning working memory: Focusing attention on specific internal representations is thought to bring the information into the focus of attention and keep them there, and conversely, when attention is shifted to other elements, the representations are moved to the activated portion of long-term memory. The distinction between the activated portion of long-term memory and the focus of attention is well illustrated by a study by Oberauer (2001). In this study, participants had to remember two memory lists, one of which was later cued as irrelevant. When recall was required immediately after the cue, reaction times were a function of the length of both lists, while they were a function of only the relevant list when the recall probe appeared one second after the cue. This suggested that the participants were able to remove the irrelevant list from the focus of attention within a second. However, the irrelevant list still had an effect on performance: When deciding whether an item was part of the relevant list, reaction times were larger when it actually came from the irrelevant list compared to completely new items. These intrusion costs were assumed to stem from the fact the irrelevant items are not completely forgotten but still reside in the activated portion of long-term memory.

Importantly, the focus of attention is assumed to be limited with respect to the number of items or chunks that can be held in memory at one point in time (Cowan, 1988, 1999). In contrast to Miller (1956) who found that immediate memory was limited to

---

3In the first and second study of the current thesis, we follow the nomenclature of Unsworth and Engle (2007) and refer to the focus of attention as primary memory and to the activated part of long-term memory as secondary memory.
2.2 Working Memory

Figure 2.3: Embedded-processes model of working memory. Figure adapted from Cowan (1988).

about 7 meaningful elements, Cowan (2001) more recently suggested that the focus of attention is limited to approximately four elements when strategies such as chunking and verbal rehearsal are explicitly prevented. Oberauer (2002) even suggested a capacity limit of a single item for the focus of attention, since in his study repeated processing of the same item was faster than switching between items. However, Oberauer’s model additionally includes the region of direct access, which is thought to be able to hold about 4 items at the same time. The limited capacity of the region of direct access is argued to stem from interference between bindings, whereas the focus of attention of the embedded-processes model is supposedly limited with respect to the number of items per se (Oberauer, 2002; Oberauer, 2013).

The question of why working memory, and ultimately the human mind, is limited is probably one of the main reasons for its popularity. Daneman and Carpenter (1980) were not only the first to show that there are considerable individual differences in WMC but also that these differences were strongly related to higher cognitive functions (in this case, reading comprehension). Following this initial finding, there has been a wealth of research showing that measures of WMC, such as the complex span tasks described
2.3 Conceptual Learning (and the Role of Prior Knowledge)

above (see p. 12), are among the best predictors of intelligence, in particular reasoning ability (e.g. Ackerman, Beier, & Boyle, 2005; Conway et al., 2011; Kane et al., 2004; Kyllonen & Christal, 1990; Oberauer et al., 2005). Although both are closely associated, the consensus seems to be that reasoning ability is still "a bit more" than WMC. In contrast to intelligence (see section 2.1, p. 9), one crucial feature of working memory is that it is derived from cognitive theories specifying the properties and limits of such an architecture. Thus, studying working memory and its capacity limit is seen as a promising way to reveal the cognitive basis of intelligence and reasoning ability. However, since there are different theories about the cognitive basis of WMC, there are also different explanations for the association between working memory and reasoning ability. The best known theory assumes that domain-general executive attention invested to maintain information, in particular in the presence of distracting and interfering information, limits working memory and drives the relation with reasoning ability (Engle, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004). In an extension of this account, the controlled retrieval from long-term memory is seen as an additional factor underlying WMC (Unsworth & Engle, 2007). As pointed out above, others emphasize storage capacity and the number of items that can be simultaneously kept in an accessible form (Cowan, 2001; Cowan et al., 2005; Halford, Wilson, & Phillips, 1998; Luck & Vogel, 1997, 2013). A more detailed description of these theoretical accounts can be found in the first and second empirical study of this thesis (see chapter 3 and 4). The goal of the first study was to compare these mechanisms with respect to reasoning ability both on the behavioral and the neural level by measuring brain activity with EEG, while the second study investigated their predictive value for conceptual learning in real classrooms. Conceptual learning and the role of prior knowledge during learning will be the topic of the next section.

2.3 Conceptual Learning (and the Role of Prior Knowledge)

The ability to divide specific instances into classes is a central aspect of human cognition (Ashby & Maddox, 2011). For example, it helps to decide whether running from or petting a particular cat-like animal is the better idea but also assists in selecting the correct solution algorithm for algebraic equations involving either addition or multiplication (e.g., \( a + a = 2 \) or \( a \cdot a = a^2 \)). Accordingly, the learning of categories and concepts
2.3 Conceptual Learning (and the Role of Prior Knowledge)

has received a lot of attention in both cognitive and educational psychology, but both fields have largely focused on different phenomena (see Goldwater & Schalk, 2016).

In cognitive psychology, a major point of interest was the characterization of how categories are represented, which has produced different theoretical views (for a summary see Goldstone & Kersten, 2003). For example, prototype accounts assume that novel instances are compared and categorized with respect to learned prototypes (Rosch, 1975), exemplar accounts assume that individual instances rather than prototypes are stored and that categorization occurs based on similarity to those instances (Kruschke, 1992) and rule-based accounts assume that logical rules are used for categorization (e.g., "if an animal has feathers and lays eggs, it is a bird"; Ashby & Maddox, 2005). Contemporary approaches, however, generally converge on the idea that a combination of representation types is required to adequately account for human category learning (e.g., Erickson & Kruschke, 1998; Kruschke, 2005).

The experimental method that was most used to investigate the acquisition of categories is the inductive classification paradigm. In this paradigm, participants typically sit in front of a computer screen, are repeatedly presented with a small, finite set of instances and must learn to correctly assign them to categories based on trial-and-error and corrective feedback. The stimuli typically differ in several well-controlled dimensions. For example, the classic problem types by Shepard, Hovland, and Jenkins (1961) vary along the three binary dimensions of color (black vs. white), size (large vs. small), and shape (square vs. triangle). This allows for the creation of category structures with different difficulty levels based on how the individual instances are assigned to categories: For example, a simple category structure simply contrasts all \textit{black} instances with all \textit{white} instances (i.e., size and shape are irrelevant), while a more complex structure contrasts \textit{black and triangular} or \textit{white and square} with \textit{white and triangular} or \textit{black and square} instances (i.e., size is irrelevant). Novel, untrained exemplars are often presented at the end of learning, which are specifically designed to draw inferences about how participants represent the category structure (e.g., is a black circle assigned to the same category as the black squares/triangles?). This strict experimental control over the learning material and situation has resulted in many insights, but has at the same limited its applicability for learning in more ecologically valid settings such as in school. One central criticism is the fact that categories in real life are never learned by themselves, i.e., in isolation from each other, but that already acquired knowledge (i.e.,
2.3 Conceptual Learning (and the Role of Prior Knowledge)

prior knowledge) strongly affects how information is perceived, interpreted, and thus learned (see Goldwater & Schalk, 2016; Murphy & Medin, 1985).

Indeed, the goal of education in general is not only to help developing an understanding of many individual concepts but a coherent and hierarchical system of knowledge involving the interrelations between concepts. For example, to understand Newton’s second law \(F = m \cdot a\), the concepts of velocity \(v = \frac{s}{t}\), acceleration \(a = \frac{v}{t}\), and the multiplicative relation of mass and acceleration must be integrated. Accordingly, the influence of prior knowledge has been a major area of interest in educational research. There is a wealth of findings demonstrating that prior knowledge is the single most important predictor of conceptual learning (e.g., Ackerman, 2007; Hambrick & Meinz, 2011; Schneider, Körkel, & Weinert, 1989; Stern, 1999, 2015; Tricot & Sweller, 2013; Weinert, Helmke, & Schneider, 1990). For example, Weinert et al. (1990) reported that mathematical achievement in 5th graders was strongly predicted by a pretest score collected one year earlier, even after controlling for differences in intelligence \(r = 0.66\); see also Stern, 2009).

However, there are also situations where prior knowledge hinders rather than promotes learning. The research on conceptual change (Carey, 2000; Vosniadou, 2008) has repeatedly shown that students are not "blank slates" when they enter the classrooms. Instead, they have acquired prior knowledge or preconceptions about many phenomena, which might interfere with learning the correct scientific concepts. For example, with respect to inertia according to Newtonian mechanics, students often wrongly assume that a continuous force must be applied to keep an object in motion due to their everyday experience that moving an object (against friction) is strenuous (Clement, 1982). Such preconceptions are very persistent and the development of the scientifically correct concepts is thought to require an extensive restructuring of conceptual knowledge. One major aim of educational research has thus been to develop instructional methods to improve conceptual learning such as comparing and contrasting cases (Gick & Holyoak, 1983; Rittle-Johnson & Star, 2007), giving self-explanations (Chi, Leeuw, Chiu, & LaVancher, 1994), and asking meta-cognitive questions (Beeth, 1998).

Furthermore, the essential role of domain-specific knowledge for superior performance has not only been highlighted in educational but also in expertise research (e.g., Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson & Towne, 2010). It has been argued that the most important factor for achieving expertise in a field is the amount of delib-
erate practice, typically years, invested in a particular domain. For example, Ericsson et al. (1993) retrospectively estimated that expert pianists had invested around 10000 hours into practice until age 20, while amateurs invested substantially less (around 2000 hours). That a broad knowledge base in the specific domain (and not innate talent or differences in general ability; see section 2.4) is the basis of the superior performance of experts has mainly been shown using the methodology introduced by Chase and Simon (1973). In their classical study, chess experts and novices were presented with either meaningful or random chess constellations and the results showed that chess experts showed better memory performance only for the meaningful but not the random constellations.

Taken together, considering prior domain-specific knowledge is essential for an adequate account of learning: Information is always interpreted and constructed with respect to what one already knows (which may both assist and hinder learning) and experts differ from novices in the amount and organization of their domain-specific knowledge. Thus, to increase the ecological validity of category learning investigated in the laboratory, prior knowledge should be considered, for example by including sequential learning phases and inducing prior knowledge via previous instruction. The advantage of the experimental approach is that it provides good control over many factors that potentially influence learning in educational settings, thus helping to gain a better understanding of the underlying cognitive processes. This is part of the reasoning that shaped the third study of the present thesis. There, we developed a category learning paradigm with two phases that allowed us to investigate the interplay of cognitive abilities and prior knowledge during conceptual learning in a well-controlled manner. The next section will provide a general overview of the impact of cognitive abilities on learning.

### 2.4 The Impact of Cognitive Abilities on Learning

Theoretically, the link between cognitive abilities such as reasoning ability and working memory and learning, i.e. the rate of improvement through practice or instruction, seems natural. Intelligence tests were specifically designed to capture interindividualse differences in learning potential (see section 2.1, p. 7) and intelligence was even defined as the "ability to learn" (e.g., Buckingham, 1921). However, early experimental studies investigating this association were rather unsuccessful in this regard. Woodrow
2.4 The Impact of Cognitive Abilities on Learning

(1938, 1946), for example, found that intelligence was not related to learning gains in simple laboratory tasks, leading him to state that the "ability to learn cannot be identified with the ability known as intelligence" (Woodrow, 1946, p. 148). The critique brought forward in the following years with respect to this conclusion has been manifold (for a summary see Lohman, 1999).

A first criticism was concerned with the fact that cognitive abilities are not expected to be related to all kinds of learning, such as classical conditioning or implicit learning (Estes & Sternberg, 1982). Similarly, the influence of cognitive abilities on learning "closed skills", for which the acquirable knowledge is relatively finite and foreseeable (e.g., typical procedural skills such as learning to type), has been suggested to be mainly limited to the beginning of practice (Ackerman, 1988, 2007). In contrast, in the case of "open skills", where no foreseeable end state of knowledge exists and one skill helps to acquire a more complex one, cognitive abilities are more likely to influence learning, even on high levels of expertise. For example, basic arithmetic skills form the basis for solving algebraic equations, which are, in turn, required to solve differential and integral equations in calculus. Or with respect to Newtonian mechanics in physics, an understanding of basic concepts such as velocity and acceleration is essential to understand the concept of force.

Another critique, although not independent of the first one, was that the tasks applied by Woodrow were too simple. For example, Woodrow (1938) related intelligence to performance in tasks requiring horizontal adding, substitution, and estimating lengths, and found an average correlation near zero ($r = 0.08$). Indeed, there is now a wealth of findings showing that cognitive abilities are predictive of learning, both in artificial laboratory tasks as well as in educational settings, when the learning material is more complex (e.g., Alexander & Smales, 1997; Lewandowsky, Yang, Newell, & Kalish, 2012; McDaniel, Cahill, Robbins, & Wiener, 2014; Primi, Ferrão, & Almeida, 2010; Voelkle, Wittmann, & Ackerman, 2006; Williams & Pearlberg, 2006). The influence of cognitive abilities is particularly evident when learning requires to go beyond the information given and when reasoning must be applied to abstract and integrate pieces of knowledge.

Conceptual learning in STEM fields is clearly one of the areas where learning is highly incremental, in the sense that one skill is a component of a more complex skill, and strongly depends on reasoning processes. In STEM education, one of the central
goals is to provide students with abstract knowledge that can be flexibly transferred to many different problems and contexts (De Corte, 2003). In particular, analogical reasoning theories have been recognized as a powerful approach to promote knowledge transfer and conceptual learning, thus bridging the gap between cognitive and educational science (Gentner, 2010). According to the structure-mapping theory (Gentner, 1983), the comparison of two exemplars (e.g., planets encircling the sun vs. electrons encircling the atom core), which might appear very different with respect to their surface features (e.g., size of the sun vs. atom core), induces structural alignment that helps to extract their common relational structure. This is then thought to allow for the transfer of knowledge from one situation (solar model) to the other (atom model). For example, using a laboratory-based category learning paradigm, Kurtz, Boukrina, and Gentner (2013) recently showed that the comparison of exemplars improved relational abstraction and transfer performance. These ideas have successfully been applied to promote conceptual learning and transfer in classrooms (e.g., Alfieri, Nokes-Malach, & Schunn, 2013; Rittle-Johnson & Star, 2007; Ziegler & Stern, 2014).

However, another important aspect that affects the potential of cognitive abilities to affect learning is the presence of prior knowledge (Lohman, 1999; Woodrow, 1946). Statistically, when the initial performance is relatively low but improves through practice or instruction, the gain will be strongly affected by the final performance, which makes the influence of cognitive abilities on learning gains more likely. In contrast, when there are already substantial individual differences in knowledge prior to the training or instruction (which is generally the case in most real-world tasks), these will be correlated with both final performance and cognitive abilities, thus essentially limiting a potential effect of cognitive abilities. Empirically, this is mirrored in findings showing substantial correlations of cognitive abilities with achievement, but less strong (or even no) associations with rate of improvement, or put differently, the loss of impact of cognitive abilities on learning when prior knowledge is controlled for (Lohman, 1999; Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2013; Primi et al., 2010; Weinert et al., 1990; Woodrow, 1946). However, the fact that prior knowledge and cognitive abilities are developmentally interwoven in most complex tasks, with cognitive abilities being invested in the acquisition of knowledge, makes it difficult to pinpoint their unique effects on learning.

Taken together, cognitive abilities are central to successful conceptual learning, but
there are several factors, such as the presence of prior knowledge, which moderate this relationship. This inspired the second and third study of this thesis in which we investigated the role of cognitive abilities for conceptual learning. In the second study, we were mainly interested in working memory as a resource for conceptual learning in an educational context. In contrast, in the third study, conceptual learning was not assessed in classrooms but in the laboratory with a newly developed category learning task.

2.5 Research Questions and Summary of Empirical Studies

In the previous sections, I have outlined that there are several open issues with respect to the interrelation between reasoning ability, WMC, and conceptual learning. The overall aim of the present thesis was to address several of those in three empirical studies drawing on issues and methods from educational psychology, cognitive/experimental psychology, and cognitive neuroscience. Specifically, the following broad research questions were formulated to tackle these issues:

1. What is the neurocognitive basis of WMC and reasoning ability?
2. What is the role of specific working memory functions for conceptual learning?
3. How do cognitive abilities and prior knowledge interplay during conceptual learning?

The first research question motivated the first study of the present thesis. Research has repeatedly shown that WMC is strongly related to intelligence, in particular to reasoning ability. The fact that working memory is a construct based on well-circumscribed cognitive theories raised expectations that studying this association will reveal the cognitive and biological basis of fluid intellectual abilities. However, there are still several candidate mechanisms discussed in the literature, including attention control, storage capacity and the controlled retrieval from long-term memory; and the findings have generally been inconclusive. Therefore, in this study, we also collected brain activity with EEG in addition to behavioral measures while participants performed a selection of working memory tasks representative for these mechanisms. In contrast to performance measures, which reflect only the end result of all processing steps (i.e., the behavioral response), neural measures allow to visualize the cognitive processes during...
their execution. This approach has the potential to uniquely inform the debate about the neurocognitive basis of WMC and reasoning ability.

The second research question was addressed in a further study that involved the same sample as the first study. However, instead of the association between WMC reasoning ability, we were interested in working memory as a resource for conceptual learning in school. The recruited participants had already taken part in a classroom study that investigated the growth in conceptual knowledge about Newtonian mechanics following an instruction encompassing 18 lessons in Swiss higher secondary schools. We expected that working memory performance would be generally predictive of conceptual learning. However, the ability for controlled retrieval of relevant information from long-term memory was assumed to be particularly important because of the central role of prior knowledge for conceptual learning. In addition, we were interested whether neural measures provide insights into physics learning over and above behavioral measures.

The third research question was mainly addressed in the third study. It is a well-known finding in educational research that prior knowledge is the main determinant of conceptual learning, which has, however, largely been neglected in the experimental research on category learning. The fact that individuals in real-world settings almost always differ in initial performance, which is typically correlated with cognitive abilities, makes it difficult to capture the specific impact of cognitive abilities and prior knowledge on conceptual learning. The motives that inspired the third study of the present thesis were thus twofold: First, we aimed to bridge the gap between experimental and educational research on category and conceptual learning by considering the influence of prior knowledge in a category learning paradigm and second, we wanted to take advantage of the control awarded by the experimental approach to investigate how cognitive abilities and prior knowledge (as well as learning strategy) interplay during conceptual learning. Accordingly, we assessed reasoning ability and WMC and designed a laboratory-based category learning paradigm that consisted of two subsequent learning phases: The first phase served to induce prior knowledge, which had to be refined in the second learning phase. This approach allowed for a fine-grained analysis of how these variables influenced conceptual learning. Each of the three empirical studies will be described in more detail in the following chapters.
References


References


References


References


References


References


References


References


32


3 The Neurocognitive Basis of Working Memory Capacity and Reasoning Ability: An EEG Study

Bruno Rütschea, Sarah Hofera, Elsbeth Sterna

aETH Zurich, Research on Learning and Instruction

Abstract

Interindividual differences in working memory capacity and reasoning ability are closely associated, but the responsible cognitive mechanisms are still a matter of debate. Several potential mechanisms are discussed in the literature, including attentional control, primary memory and retrieval from secondary memory. In an attempt to better understand this association, we examined these mechanisms not only from a behavioral but also from a neural perspective. Particularly, we measured oscillatory activity in the theta and alpha frequency band with EEG and examined the relationship between reasoning ability and brain activity during working memory tasks representative for these mechanisms mentioned above. Statistical analysis revealed that theta activity was elevated in conditions with increased need for control mechanisms, which is in line with its supposed role for human working memory. There was also evidence for the neural efficiency phenomenon, with higher reasoning ability being associated with generally increased alpha activity. Furthermore, we found that brain activity in the alpha band could differentiate between participants with higher and lower reasoning ability during all suggested cognitive mechanisms. This confirms recent studies that demonstrated that multiple mechanisms rather than a single mechanism are needed to explain the neurocognitive basis of working memory capacity and reasoning ability. The present study further highlights the potential of neuroimaging methods to inform psychological theory.
3.1 Introduction

*Keywords*: working memory capacity; reasoning ability; electroencephalography; attention control; primary memory; secondary memory retrieval

*Author contributions*: BR was involved in the study design, acquisition of data, interpretation of data, and writing of the first draft of the manuscript. SH contributed to the acquisition of the data and critically revised the manuscript. ES contributed to the study design and the critical revision of the manuscript.

3.1 Introduction

Intelligence is among the psychological constructs with the highest predictive value of real-world performance in school and profession (Neisser et al., 1996; Schmidt & Hunter, 1998). Although the value of intelligence as a construct is undeniable, it has been difficult to describe the underlying cognitive processes and mechanisms (Oberauer, Schulze, Wilhelm, & Süß, 2005). This partly arises from the fact that intelligence is measured as the common variance across a diverse set of tasks that have not been constructed with an explicit theoretical concept in mind. For that reason, the study of working memory and its capacity has raised hope to gain a better understanding of intelligence, in particular of reasoning ability. In the present study, we investigated the neurocognitive basis of working memory capacity (WMC) and reasoning ability by complementing behavioral data with neural data. This provided an additional perspective that has the potential to shed a different light on and thus uniquely inform this relationship.

3.1.1 The Cognitive Basis of Working Memory Capacity

Working memory is generally conceptualized as a capacity-limited system that allows individuals to maintain, manipulate and access information in service of currently relevant goals (Baddeley, 2012; Baddeley & Hitch, 1974; Cowan, 1999). Fluid intelligence or reasoning ability is a core component in many contemporary theories of human intelligence, referring to the capacity to solve novel problems without drawing on previously acquired knowledge (Cattell, 1963, 1971). Research has generally shown that interindividual differences in WMC and reasoning ability are strongly related and share at least
3.1 Introduction

50% of variance (e.g. Kane et al., 2004; Oberauer et al., 2005). The two constructs are thus not isomorphic, although not all empirical findings are in agreement (Ackerman, Beier, & Boyle, 2005; Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Colom, Abad, Rebollo, & Chun Shih, 2005; Kyllonen & Christal, 1990). The close relation between WMC and reasoning ability has been recognized as a unique opportunity to investigate the mechanisms underlying intelligent information processing and to ultimately better understand intelligence (Oberauer et al., 2005). To this date, however, there is considerable disagreement about the nature of WMC and, in turn, about its association with reasoning ability. The most discussed mechanisms in the literature are attention control, primary memory and retrieval from secondary memory, which will be described in more detail in the following.

According to the executive attention account of WMC, working memory is mainly limited by attention control, i.e. the ability to maintain task-relevant information during distraction or interference from task-irrelevant information (Engle, 2002; Engle, Tuholski, Laughlin, & Conway, 1999). Support for this assumption comes from a strand of research showing that individuals with higher compared to lower WMC perform better in tasks requiring to overcome or inhibit a dominant, habitual response (for an overview see Engle, 2002). For example, the Stroop task requires participants to name the ink color in which a color word is printed. The word and ink color can either be congruent (i.e., matching) or incongruent (i.e., not matching). In the incongruent trials, the prepotent response (i.e., reading the word) must be inhibited to give the correct response (i.e., name the ink color). In line with the executive attention account, higher WMC individuals performed better than lower WMC individuals only when there was a bias toward habitual responding (i.e., when 75% of the trials were congruent but not when 50% of the trials were congruent; Kane & Engle, 2003).

In contrast, another theoretical perspective stresses the importance of storage capacity (Cowan, 2001; Cowan et al., 2005; Luck & Vogel, 1997). Following Unsworth and Engle (2007), we use "primary memory" to refer to this store, while acknowledging that the terms "focus of attention" (Cowan, 1999) or "region of direct access" (Oberauer, 2002) describe similar concepts. Generally, the function of primary memory is to maintain a set of representations for active processing by means of the continued allocation of attention (Unsworth & Engle, 2007). Its capacity is thought to be restricted to about 4 items, although people differ in the amount of information they are able to maintain.
3.1 Introduction

(Cowan, 2001; Luck & Vogel, 1997). The visual arrays task is one prototypical task used to measure the capacity of primary memory (Luck & Vogel, 1997). In this task, an array of objects (e.g., colored squares) is briefly presented. A second array is shown after a blank inter-stimulus interval and the participants are required to indicate whether the two arrays, which differ in one object or not at all, are identical. Typically, indicating that the capacity limit has been reached, performance strongly deteriorates when more than 3 to 4 objects are to be maintained. Vogel and Machizawa (2004) even showed that brain activity increases with array size but reaches an asymptote for arrays that meet or exceed the capacity limit.

Because in a capacity-limited system some information is inevitably displaced during cognitive performance, there must be a way to retrieve previously held information. According to the dual-component model by Unsworth and Engle (2007), information removed from primary memory remains in secondary memory. Secondary memory is thought to consist of long-term memory representations that are not activated enough to be in primary memory and thus conscious awareness. Therefore, recall from secondary memory is speculated to involve a cue-dependent search process to increase the activation of representations and bring them back into primary memory. Crucially, higher WMC is assumed to be reflected in the ability to narrowly constrain memory search by means of more specific contextual retrieval cues, which facilitates the targeting of relevant information among interfering irrelevant information. In contrast, lower WMC participants are expected to produce less specific retrieval cues, which in turn generates more irrelevant retrieval candidates. The displacement from primary to secondary memory either occurs because of new incoming information (i.e., when remembering long lists) or when attention is moved away from maintaining the items (e.g., due to a distracting intermediate task).

Related to the issue about the mechanisms involved in WMC is the question why updating tasks are related to reasoning ability. Generally, updating tasks require the continuous monitoring and updating of working memory contents (see e.g. Oberauer, Süss, Schulze, & Wittman, 2000). For example, the n-back task, which is one of the most commonly used updating tasks, requires participants to indicate whether the currently presented item in a continuous stream of items matches the one presented \( n \) steps before. According to Miyake et al. (2000), updating is an executive function that is separable from response inhibition (i.e., as required in the Stroop or Flanker task). Al-
3.1 Introduction

though updating tasks correlate well with both other working memory tasks and reasoning ability (Chuderski & Necka, 2012; Friedman et al., 2006; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009; Wilhelm, Hildebrandt, & Oberauer, 2013), the reason is unclear. Recent evidence suggests that it might be due to general maintenance and processing demands that are in common with other working memory tasks and not specifically due to the requirement to update the contents of working memory, i.e., the attention control component of the task (Ecker, Lewandowsky, Oberauer, & Chee, 2010; Wilhelm et al., 2013).

Taken together, different explanations for the close association between individual differences in WMC and reasoning ability have been proposed. Over the last years several studies have been conducted to compare the different explanations within regression or structural equation modeling (SEM) frameworks. The findings, however, were quite heterogeneous: There is research that found incremental validity of attention control measures (e.g., Cowan, Fristoe, Elliott, Brunner, & Saults, 2006; Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth & Spillers, 2010), while others downplay the role of attention control, with some specifically favoring primary memory accounts (e.g. Chuderski, Taraday, Necka, & Smolen, 2012; Colom et al., 2008; Keye, Wilhelm, Oberauer, & Van Ravenzwaaij, 2009). A similar debate has been concerned with whether retrieval from secondary memory explains unique variance over and above the other mechanisms (e.g., Mogle, Lovett, Stawski, & Sliwinski, 2008; Shipstead et al., 2014; Unsworth, 2010; Wilhelm et al., 2013). Although the relative contributions are unresolved, recently, there have been an increasing number of publications arguing for multiple rather than single mechanisms underlying the relationship between WMC and reasoning ability (Conway, Getz, Macnamara, & Engel de Abreu, 2011; Conway & Kovacs, 2015; Cowan et al., 2006; Shipstead et al., 2014; Unsworth & Engle, 2007; Unsworth & Spillers, 2010). In the present study, we wanted to shed light on these issues by including measures of neural activity in addition to performance measures.

3.1.2 The Neural Correlates of Working Memory Capacity and Reasoning Ability

The identification of the neural correlates of reasoning ability and working memory has received much attention in neuroscience. Two hypotheses are particularly prevalent
3.1 Introduction

with respect to reasoning ability or intelligence: The Parieto-Frontal Integration Theory (P-FIT) and the neural efficiency hypothesis. On the one hand, the P-FIT relates intelligence to individual differences in a network of brain areas, especially in frontal and parietal areas (Jung & Haier, 2007). Research suggests that frontal areas are more involved in attention control and parietal areas in storage processes, although both areas closely work together even for short-term maintenance (Jonides et al., 2008; Ruchkin, Grafman, Cameron, & Berndt, 2003; Sarnthein, Petsche, Rappelsberger, Shaw, & von Stein, 1998; Wager & Smith, 2003). The close association between working memory and intelligence is also apparent on the neural level, as both seem to rely on a similar network of brain areas (Jonides et al., 2008; Kane & Engle, 2002).

On the other hand, the neural efficiency hypothesis associates intelligence with efficient brain functioning, with more intelligent participants generally being characterized by less brain activation during cognitive performance (Haier et al., 1988; Neubauer & Fink, 2009; Nussbaumer, Grabner, & Stern, 2015). Most research within the neural efficiency framework has measured oscillatory activity in the alpha frequency band (~8-12 Hz) with electroencephalography (EEG). It was found that more intelligent participants show larger absolute alpha power or, when comparing alpha power during cognitive processing to a resting period, smaller decreases in alpha power (Neubauer & Fink, 2009). Classically, increased alpha activity has been interpreted to reflect a state of reduced information processing (Pfurtscheller & Lopes da Silva, 1999). More recent accounts relate synchronized alpha activity to functional inhibition of brain areas, with low activity indicating engagement and high activity indicating disengagement of specific brain areas (Jensen & Mazaheri, 2010; Klimesch, Sauseng, & Hanslmayr, 2007). Thus, the general positive relationship between intelligence and alpha activity seems to suggest that intelligence is related to the disuse of task-irrelevant areas and a more focused use of task-relevant areas (Haier et al., 1988; Neubauer & Fink, 2009).

With respect to EEG oscillations and working memory, there have been numerous reports of the close association of theta activity (~4-6 Hz) with working memory processes, in particular at fronto-medial channels (Bastiaansen & Hagoort, 2003; Klimesch, 1999; Roux & Uhlhaas, 2014; Sauseng, Griesmayr, Freunberger, & Klimesch, 2010). For example, theta activity has been found to increase with memory load (Gevins, Smith, McEvoy, & Yu, 1997; Jensen, Gelfand, Kounios, & Lisman, 2002; Sauseng, Hoppe, Klimesch, Gerloff, & Hummel, 2007), when manipulating compared to main-
3.1 Introduction

taining information (Griesmayr, Gruber, Klimesch, & Sauseng, 2010), or when solving incongruent compared to congruent trials in a Stroop task (Hanslmayr et al., 2008). Although the function of theta activity during working memory is still unresolved, it has been suggested to be part of a control system that allocates and coordinates cognitive resources (Mitchell, McNaughton, Flanagan, & Kirk, 2008; Sauseng et al., 2007; Sauseng et al., 2010). This corresponds with findings showing that the source of the fronto-medial theta activity lies in the anterior cingulate cortex (ACC; Hanslmayr et al., 2008; Onton, Delorme, & Makeig, 2005), which has been implicated in conflict monitoring and the engagement of cognitive control (Botvinick, Cohen, & Carter, 2004). However, the association with working memory is not specific to the theta band, as there are also several reports of increased alpha activity during working memory retention in the modified Sternberg and visual arrays tasks (e.g. Gevins et al., 1997; Jensen et al., 2002; Palva, Monto, Kulashekhar, & Palva, 2010; Palva, Kulashekhar, Hämäläinen, & Palva, 2011; Scheeringa et al., 2009). This is, though, thought to reflect the inhibition of task-irrelevant brain regions, in contrast to the aforementioned role of theta for control processes in working memory (Klimesch et al., 2007; Roux & Uhlhaas, 2014).

3.1.3 The Current Study

Several cognitive mechanisms are discussed to account for the close association between WMC and reasoning ability, including interindividual differences in attention control, primary memory and retrieval from secondary memory. In an attempt to better understand this association, the present study collected EEG during working memory tasks that strongly involve these mechanisms. Specifically, we investigated the relationship between reasoning ability, as measured by Raven’s Advanced Progressive Matrices (RAPM; Raven, Raven, & Cout, 2003), and oscillatory EEG activity in four different working memory tasks: First, as a measure of attention control, we used a color-word Stroop task, which included trials where the ink color and color word were either congruent (i.e., matching) or incongruent (i.e., not matching). Second, as a measure of primary memory, we used a visual arrays task with trials containing 2, 4 and 6 squares. Third, to measure retrieval from secondary memory, we used a newly developed secondary memory task. The task involved maintaining 4 locations in a 4×4 grid, performing an unrelated secondary task and then recalling the positions. Fourth, arguably
3.1 Introduction

measuring both primary memory and attention control, we used an updating task. The participants had to continuously maintain and update 3 locations in a 4×4 grid based on arrows pointing either left, right, up or down. The administration of EEG allowed us to "zoom in" and visualize the cognitive mechanisms during performance (Stern & Schneider, 2010), instead of drawing inferences about them solely based on patterns of performance across different tasks. The inclusion of the updating task, specifically, allowed us to test whether there are differences in brain activity in relation to reasoning ability during updating (attention control) that go over and above differences present during maintenance without updating (primary memory). Overall, this interindividual difference approach including neural data will help to clarify the specific cognitive processes involved in WMC and reasoning ability.

We expected that behavioral performance would be lower in incongruent compared to congruent trials in the Stroop task (Hanslmayr et al., 2008) and would generally decrease with increasing memory load (Jensen et al., 2002; Luck & Vogel, 1997). Furthermore, higher compared to lower reasoning ability was assumed to result in better performance across tasks. The expectations with respect to oscillatory theta activity were derived from previous studies using similar tasks: Theta activity was hypothesized to be increased at fronto-medial channels in incongruent compared to congruent trials in the Stroop task (Hanslmayr et al., 2008), in conditions with higher compared to lower memory load in the visual arrays and secondary memory task (e.g., Gevins et al., 1997; Palva et al., 2010), and during updating compared to maintenance in the updating task (Griesmayr et al., 2010). Furthermore, in line with prior research, we expected alpha activity to decrease in conditions with higher difficulty (i.e., incongruent compared to congruent trials in the Stroop task, updating compared to maintenance in the updating task; Gevins et al., 1997) and to exhibit a memory-load dependent increase in the visual arrays and secondary memory tasks (e.g. Jensen et al., 2002; Palva et al., 2010; Palva et al., 2011). However, in the visual arrays task, we generally expected differences in brain activity only between loads 4 and 2 but not between loads 6 and 4, because the argued capacity limit is reached at about load 4 (see also Vogel & Machizawa, 2004).

One distinguishing feature of the present study is that we also adopted an interindividual differences approach to investigate how brain activity differs with respect to reasoning ability. In accordance with the neural efficiency hypothesis (Neubauer & Fink, 2009), we assumed that reasoning ability would be characterized by generally increased
alpha activity in all tasks. However, particularly relevant with respect to the overall goal of the present study, was how reasoning ability would relate to the adaptation of brain activity to the conditions within each task, since this would directly shed light on the specific cognitive mechanisms involved in WMC. Specifically, following recent findings arguing for a multi-mechanism view of WMC, we expected to find reasoning ability-related differences in brain activity during attention control, primary memory and retrieval from secondary memory (e.g., Conway et al., 2011; Conway & Kovacs, 2015; Shipstead et al., 2014). These differences might be particularly pronounced in the alpha band due to its close association with intelligence. Overall, the present study is, to the best of our knowledge, the first to use neuroimaging to compare different theories concerning the nature of WMC. This approach will provide insights that cannot be gained by behavioral measures alone and thus help to increase the understanding of the neurocognitive basis of WMC and reasoning ability.

3.2 Method

3.2.1 Sample

We recruited 39 students (27 females; $M_{age} = 15.87$, $SD_{age} = 1.03$) attending Swiss higher secondary school (Gymnasium) from a classroom study that investigated whether conceptual learning can be enhanced via cognitively activating instruction (Hofer, Schuhmacher, Rubin, & Stern, 2017). From this study, we adopted the scores of the RAPM (Raven et al., 2003) and a learning test (hofer_2017_the_test_of_basic_mechanics). We discarded 9 participants that had missing values for either measure to have comparable samples in this study and a related study that focused on predicting learning gains (Rütsche, Hofer, & Stern, 2017). The final sample consisted of 30 participants (21 females; $M_{age} = 16$, $SD_{age} = 0.91$).

The representativeness of the final sample to the overall sample was assessed with a bootstrap procedure. The R package boot (Angelo & Brian, 2016) was used to draw 3000 random samples (N = 30) from the overall sample (N = 285). 95%-confidence intervals of the mean and standard deviation for RAPM and age as well as the gender composition were computed (males coded as 0; females coded as 1). We then checked whether the values of the final sample lay within or outside of the bootstrapped con-
3.2 Method

Confidence intervals. The sample was representative with respect to RAPM scores ($M = 27.6, 95\%-\text{CI} [25.5 28.9]; SD = 4.82, 95\%-\text{CI} [3.16 5.36]$), age ($M = 16, 95\%-\text{CI} [15.6 16.2]; SD = 0.91, 95\%-\text{CI} [0.693 1.234]$) and gender ($M = 0.3, 95\%-\text{CI} [0.28 0.61]$). The study was approved by the local ethics committee of the Swiss Federal Institute of Technology, Zurich, Switzerland and all parents of the participants gave written informed consent. The participation was reimbursed with 100 CHF (about 100 USD; plus travel expenses when traveling to the EEG laboratory with public transportation).

3.2.2 Procedure

At the follow-up session of the classroom study (Hofer et al., 2017), the students were asked if they were interested in taking part in a further study about the neural basis of cognitive abilities. They could either write down their email address on a list that was distributed in class or contact the first author directly via email. Individual sessions were then arranged via email.

The experimental session lasted about 3.5 hours overall. Upon arrival of the participants, we installed the EEG and seated them in an electromagnetically shielded cabin to reduce electromagnetic interference during EEG recording. After a short briefing about the EEG measurement, the participants performed the following working memory tasks: operation span, updating, visual arrays, secondary memory, and Stroop task. The order of the tasks was held constant across participants. Self-timed pauses were included between as well as within tasks. After EEG recording, the participants were asked to solve the subtests “verbal analogies”, “number series”, “figure selection” and “matrices” from the intelligence test I-S-T 2000 R (Liepmann, Beauducel, Brocke, & Amthauer, 2007). All participants were thoroughly debriefed and compensated for their participation at the end of the session.

3.2.3 Measures

The main measures used in the present study were the RAPM (Raven et al., 2003), the Stroop, visual arrays, secondary memory and updating tasks. The working memory tasks were all programmed with PsychoPy v1.81.00 (Peirce, 2007, 2009) and the same randomly determined sequence of stimuli was presented to all participants. These measures will be discussed in more detail in the following sections. Several other mea-
3.2 Method

sures, although collected, were not used for the statistical analyses. The operation span adopted from Lewandowsky, Oberauer, Yang, and Ecker (2010) simply served to habituate the participants to the experimental situation and thus had no specific hypothesis associated with it. Furthermore, we used RAPM instead of the I-S-T 2000 R subtests as a measure of reasoning ability, because the latter showed surprisingly low and/or inconsistent correlations with the working memory tasks (ranging from -0.13 to 0.36) as well as with each other (ranging from 0.03 to 0.5). Using the RAPM has the additional advantage that the findings are more comparable to other research in this area (Neubauer & Fink, 2009).

Reasoning ability

The RAPM (Raven et al., 2003) was used as a standardized measure of fluid intelligence and reasoning ability. The test was conducted in classrooms according to the test manual instructions. The objective of the RAPM is to abstract relational rules underlying a 3×3 matrix of geometrical figures and fill in a missing piece with 1 of 8 response alternatives. Set I (12 problems) was used as a training set and presented without time constraint. Set II (36 problems) was administered with a standard time constraint of 40 min. The mean-centered sum score of Set II was used as the measure of reasoning ability in all statistical analyses.

Stroop task

In the Stroop task, the German color words ROT (RED), GRÜN (GREEN), or BLAU (BLUE) were presented on a gray background in either red, green or blue ink (see Figure 3.1). The word and ink color were either congruent (i.e., RED in red ink, GREEN in green ink, and BLUE in blue ink) or incongruent (i.e., RED in blue or green ink, GREEN in red and blue ink, BLUE in red or green ink). The participants were required to give a response with respect to the ink color irrespective of the color word. Each stimulus was shown until the participants responded via colored buttons on a Serial Response Box (Psychology Software Tools, Sharpsburg, PA) or until a time-out of 1.25s was reached. The intertrial interval showed a fixation point for a randomly determined duration between 1.25 and 1.75s.

Twelve practice trials with performance feedback were presented at the beginning of
3.2 Method

the task. Afterwards, 270 test trials composed of 60% congruent and 40% incongruent trials were shown in random order. Previous studies have shown that a larger proportion of congruent trials increases interference effects and attentional demands of the task, because a bias toward habitual responding (i.e., answering based on word information) is developed, which has to be overcome in incongruent trials (e.g., Kane & Engle, 2003; Tillman & Wiens, 2011). The number of trials for all combinations of word and ink color within the congruent and incongruent trials was constant: Each color word was presented with the matching ink color 36 times (congruent trials) and with each non-matching color 6 times (incongruent trials). We computed the solution rates across all trials and separately for congruent and incongruent trials.

Figure 3.1: Trial structure of the Stroop task.

Visual Arrays task

In the visual arrays task (adapted from Cowan et al., 2005; Vogel & Machizawa, 2004), an array of 2, 4, or 6 colored squares was presented for 0.1s in each trial (see Figure 3.2). A second array was presented after a blank inter-stimulus interval shown for 0.9s. This array was either identical to the first one or differed in the color of a single square. Additionally, one square was circled and the participants were informed that the two arrays either differed in the color of the circled square or not at all. The participants
3.2 Method

indicated in each trial whether the color in the second array changed or not via two buttons on the button box (left button: different, right button: same). The trial ended after the response or when the time-out of 2s was reached. Afterwards, the intertrial interval showed a fixation point for a randomly determined duration between 1.25s and 1.75s.

After 12 practice trials with performance feedback, 240 test trials without feedback were presented. The task consisted of an equal number of trials for each load (i.e., 2, 4, and 6 squares), involving a color change in half of the trials. Based on an estimated viewing distance of 50 cm, the squares had a size of 0.65° and were presented centered on the screen within a 7.3°×7.3° region. The locations of the individual squares were randomly determined, although they were required to be at least 2° apart from each other and from the fixation point at the center of the screen. The color of each square was randomly assigned to be either red, green, blue, yellow, violet or white. Each color could not appear more than twice per array. We computed the solution rates across all trials and separately for each load.

![Figure 3.2: Trial structure of the visual arrays task.](image)
3.2 Method

Secondary Memory task

The secondary memory task consisted of four consecutive phases: a maintenance, calculation, recall, and response phase (see Figure 3.3). In the maintenance phase, the participants were required to remember the order of 4 squares which were sequentially presented within a 4×4 grid (field of view: 6.3°×6.3°, individual cell size: 1.575°). Each square was displayed for 0.5s followed by an interstimulus interval showing an empty grid for 1s. The calculation phase involved solving four simple arithmetic equations. The equations were adopted from the working memory test battery developed by Lewandowsky et al. (2010). The participants indicated for each equation whether it was correct (left mouse button) or not (right mouse button) within a time-window of 3s. The recall phase was indicated by a question mark displayed below the grid, upon which the participants were instructed to recall the maintained square positions and then press the left mouse button. This initiated the response phase in which the participants selected the maintained positions in serial order by clicking in the respective cells. We opted for a list-wise recall as a pilot study indicated that participants typically recalled the whole list even when cued for individual items. At the end of the trial, a fixation point was shown for a randomly determined duration between 2.75s and 3.25s.

The participants were familiarized with the task with four practice trials that included performance feedback and then solved 35 test trials without feedback. Two successive square positions could not be in adjacent cells to reduce the likelihood of patterns that facilitate maintenance. We computed the solution rates across all trials and separately for each position (i.e., the proportion of correctly recalled items for positions 1 to 4).

Updating task

The updating task (adapted from Chen & Li, 2007) consisted of three consecutive phases: a maintenance, updating, and recall phase (see Figure 3.4). In the maintenance phase, the positions of one red, one green, and one blue square presented within a 4×4 grid had to be remembered (field of view: 6.3°×6.3°, individual cell size: 1.575°). Each colored square was displayed for 0.5s, followed by an interstimulus interval showing an empty grid for 1s. In the updating phase, 4, 5, or 6 colored arrows were presented in succession at the center of the grid. The orientation of the colored arrows (either left,
3.2 Method

Figure 3.3: Trial structure of the secondary memory task.

right, up, or down) indicated the direction the square with the corresponding color had to be moved, i.e. updated. The timings were the same as in the maintenance phase. The end of the updating phase and the beginning of the recall phase was indicated by a question mark displayed below the grid for 1s. Red, green, and blue squares were now shown one by one under the grid in random order, prompting the participants to select the last position of each color by clicking in the respective cell. The intertrial interval showed a fixation point for a randomly determined duration between 2.75s and 3.25s.

The participants practiced the task with 3 trials with performance feedback and then performed 24 test trials without feedback. The task consisted of an equal number of trials with 4, 5 or 6 arrows (i.e., updating steps). The variation in updating steps ensured that participants were not able to guess the end of the updating sequence. Within a trial, each color was updated at least once and the same color was not updated twice in a row. Furthermore, a square was not allowed to be moved to its previous position to ensure that updating is indeed required at each step. We computed the solution rates across all trials and separately for trials with 4, 5, or 6 updating steps.
3.2 Method

3.2.4 EEG Recording and Preprocessing

EEG data were recorded with a BioSemi ActiveTwo system (BioSemi, Amsterdam, The Netherlands) at a sampling rate of 1024Hz. 64 active electrodes were mounted in elastic BioSemi head caps with electrode positions according to the extended 10–20 system (see Figure 3.5). The feedback loop between the Common Mode Sense active and Driven Right Leg passive electrode was used as the online reference. We kept electrode direct current offsets consistently below 40 mV and if possible below 20 mV.

Data preprocessing was performed in FieldTrip (version 02-10-2016 Oostenveld, Fries, Maris, & Schoffelen, 2011) run under 64 bit MATLAB 2016a. A Butterworth high-pass filter at 0.5 Hz (filter order: 5, two passes) and a band-stop filter between 48 and 52 Hz (and its harmonic between 98 and 102 Hz) were applied to remove DC drift and power line noise. Data were downsampled to 256 Hz. Afterwards, artifactual segments and bad channels were first detected with the automatic procedures implemented in FieldTrip and complemented by visual inspection. Independent Component Analysis (ICA) was performed on the cleaned data separately for each subject and task (Delorme & Makeig, 2004). Independent components reflecting eye blinks and vertical/horizontal eye movements were rejected based on topography and temporal progression. Spheri-
3.2 Method

Figure 3.5: Schematic of the EEG layout. F = frontal, T = temporal, C = central, P = parietal, O = occipital, p = polar, z = zeros (at midline)

cal spline interpolation was applied to restore the rejected channels and the data were referenced to an average reference.

Next, we extracted artifact-free epochs from the continuous data and computed trial-wise time-frequency representations for frequencies 3-12 Hz (in integer steps). A fast Fourier transform approach with a single Hanning taper and an adaptive time window of 3 cycles for each frequency ($\Delta T = \frac{3}{f}$) that moved in steps of 50ms across the epochs was applied (Osipova et al., 2006). We averaged across the time from 0.4s to 0.7s after the appearance of the color word in the Stroop task (Hanslmayr et al., 2008) and from -0.6s to 0s before the appearance of the second array in the visual arrays task (Palva et al., 2010; Vogel & Machizawa, 2004). In the updating task, we averaged across time from 0.2s to 1.5s after the presentation of the last item in the maintenance phase (i.e., the last item without updating) and after each item in the updating phase. This allowed us to compare brain activity during maintenance and updating without load-related confounds. In the secondary memory task, we averaged across time from 0.2s to 1.5s after each item in the maintenance phase and from 0.2s until 0.25s before the button press in the recall phase. We then created an average topography in the theta (3-6 Hz) and the alpha band (8-12 Hz) separately for each task and condition and computed activation...
3.2 Method

differences between conditions (e.g., difference in theta/alpha activity between congruent and incongruent trials in Stroop task). No baseline correction was applied, since this would make the interpretation of some effects, in particular those with respect to the differences between conditions, problematic.

3.2.5 Statistical Analyses

Behavioral data was analyzed with linear mixed effects regression using the \textit{lme4} package (Bates, Mächler, Bolker, & Walker, 2015) in the R environment (R Development Core Team, 2016). For each task, a model predicting the performance based on reasoning ability (i.e., the mean-centered RAPM Set II score), condition (using sum contrasts) and their interaction was computed. A random intercept for participants was included to account for repeated measurements. Statistical significance of the fixed effects was assessed with F tests and type 3 sums of squares using the \textit{afex} package (Singmann, Bolker, Westfall, & Aust, 2016). The package \textit{lsmeans} (Lenth, 2016) was used to perform pairwise post-hoc comparisons based on least-squares means. Degrees of freedom were approximated with the Kenward-Roger method (Højsgaard, 2016) and multiple comparisons were accounted for with Tukey’s method.

With respect to the EEG data, we generally included only trials or items that were correctly solved. However, because the task structure did not allow to identify brain activity selectively related to correct items in the secondary memory task, we analyzed the complete recall activity. Instead, we included the overall performance in the task as a covariate in the statistical analyses to control for potential effects on brain activity. The statistical significance of the EEG data was established by means of cluster-based permutation tests implemented in FieldTrip (Maris, 2012; Maris, Schoffelen, & Fries, 2007; Oostenveld et al., 2011). This approach controls the family wise error rate (i.e., the probability of type 1 errors) by clustering samples with the same effect based on temporal, spectral, and/or spatial adjacency (i.e., neighboring channels).

First, sample-wise \( t \)-values were calculated separately for each channel and frequency band in several ways: a) The difference between two conditions (i.e., difference between incongruent and congruent trials in the Stroop task; difference in neighboring loads in the visual arrays and secondary memory task) or the difference in phases in the updating task (i.e., updating minus maintenance phase) was evaluated with dependent-
3.3 Results

3.3.1 Behavioral Data

Table 3.1 shows the descriptive statistics and Table 3.2 the correlations between all behavioral measures.

In the Stroop task, we found a significant effect of condition ($F(1, 28) = 46.95, p < 0.001$), reflecting that congruent trials were solved more correctly than incongruent trials. Participants with higher compared to lower reasoning ability performed better overall ($F(1, 28) = 6.17, p = 0.02$), but reasoning ability did not interact with congruency ($F(1, 28) = 2.89, p = 0.10$).

In the visual arrays task, performance significantly differed between different loads ($F(2, 56) = 95.37, p < 0.001$): Trials with load 2 were easier than those with load 4.
3.3 Results

Table 3.1: Descriptive statistics of the behavioral data.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RAPM</strong></td>
<td>27.63</td>
<td>4.82</td>
<td>19.00</td>
<td>34.00</td>
</tr>
<tr>
<td><strong>Stroop</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>0.96</td>
<td>0.03</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>congruent</td>
<td>0.98</td>
<td>0.02</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td>incongruent</td>
<td>0.93</td>
<td>0.05</td>
<td>0.81</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Visual arrays</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>0.90</td>
<td>0.04</td>
<td>0.82</td>
<td>0.98</td>
</tr>
<tr>
<td>load 2</td>
<td>0.96</td>
<td>0.03</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td>load 4</td>
<td>0.92</td>
<td>0.04</td>
<td>0.83</td>
<td>0.99</td>
</tr>
<tr>
<td>load 6</td>
<td>0.84</td>
<td>0.06</td>
<td>0.73</td>
<td>0.98</td>
</tr>
<tr>
<td><strong>Secondary memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>0.66</td>
<td>0.15</td>
<td>0.31</td>
<td>0.89</td>
</tr>
<tr>
<td>load 1</td>
<td>0.68</td>
<td>0.15</td>
<td>0.37</td>
<td>0.89</td>
</tr>
<tr>
<td>load 2</td>
<td>0.68</td>
<td>0.17</td>
<td>0.31</td>
<td>0.94</td>
</tr>
<tr>
<td>load 3</td>
<td>0.65</td>
<td>0.16</td>
<td>0.34</td>
<td>0.91</td>
</tr>
<tr>
<td>load 4</td>
<td>0.65</td>
<td>0.17</td>
<td>0.20</td>
<td>0.94</td>
</tr>
<tr>
<td><strong>Updating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>0.80</td>
<td>0.16</td>
<td>0.25</td>
<td>1.00</td>
</tr>
<tr>
<td>4 steps</td>
<td>0.82</td>
<td>0.17</td>
<td>0.21</td>
<td>1.00</td>
</tr>
<tr>
<td>5 steps</td>
<td>0.81</td>
<td>0.17</td>
<td>0.29</td>
<td>1.00</td>
</tr>
<tr>
<td>6 steps</td>
<td>0.76</td>
<td>0.19</td>
<td>0.25</td>
<td>1.00</td>
</tr>
</tbody>
</table>

RAPM = Raven’s Advanced Progressive Matrices.

\( (M_{diff} = 0.04, t(56) = 4.59, SE = 0.01, p < 0.001) \) and 6 \( (M_{diff} = 0.12, t(56) = 13.57, SE = 0.01, p < 0.001) \); trials with load 4 were easier than those with 6 \( (M_{diff} = 0.08, t(56) = 8.99, SE = 0.01, p < 0.001) \). Reasoning ability was positively associated with overall performance \( (F(1, 28) = 6.67, p = 0.02) \), but there was no interaction between reasoning ability and load \( (F(2, 56) = 0.61, p = 0.55) \).

In the secondary memory task, performance did not depend on load \( (F(3, 84) = 2.58, p = 0.06) \), although there was a tendency that earlier items were better remembered than later items (see Table 3.1). Participants with higher reasoning ability made overall less errors \( (F(1, 28) = 5.26, p = 0.03) \). We found no interaction between reasoning ability
3.3 Results

and load \((F(3, 84) = 0.86, p = 0.47)\).

In the updating task, performance was affected by the number of updating steps \((F(2, 56) = 3.47, p = 0.04)\), but none of the post-hoc contrasts survived the correction for multiple comparisons (4 vs. 5: \(M_{diff} = 0.004, SE = 0.03, t(56) = 0.17, p = 0.99\); 4 vs. 6: \(M_{diff} = 0.06, SE = 0.03, t(56) = 2.36, p = 0.06\); 5 vs. 6: \(M_{diff} = 0.06, SE = 0.03, t(56) = 2.2, p = 0.09)\). Again, reasoning ability was related to overall good performance \((F(1, 28) = 10.54, p = 0.00)\), but did not interact with the number of updating steps \((F(2, 56) = 0.20, p = 0.82)\).

Table 3.2: Correlation matrix of the behavioral data.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 RAPM</td>
<td>0.43*</td>
<td>0.44*</td>
<td>0.40*</td>
<td>0.52*</td>
<td></td>
</tr>
<tr>
<td>2 Stroop</td>
<td>0.43*</td>
<td>0.47*</td>
<td>0.23</td>
<td>0.38*</td>
<td></td>
</tr>
<tr>
<td>3 Visual arrays</td>
<td>0.44*</td>
<td>0.47*</td>
<td>0.35</td>
<td>0.51*</td>
<td></td>
</tr>
<tr>
<td>4 Secondary memory</td>
<td>0.40*</td>
<td>0.23</td>
<td>0.35</td>
<td>0.58*</td>
<td></td>
</tr>
<tr>
<td>5 Updating</td>
<td>0.52*</td>
<td>0.38*</td>
<td>0.51*</td>
<td>0.58*</td>
<td></td>
</tr>
</tbody>
</table>

RAPM = Raven’s Advanced Progressive Matrices; *p < 0.05.

3.3.2 Neural Data

The results of the cluster-based permutation tests for the Stroop task are depicted in Figure 3.6. We found that incongruent trials were characterized by higher theta activity than congruent trials at frontal to central channels in both hemispheres, with the largest difference showing at fronto-medial channels \((p < 0.01)\). Incongruent trials compared to congruent trials were also accompanied by less alpha activity at parieto-occipital channels \((p < 0.01)\). With respect to reasoning ability, a significant difference in the alpha band \((p = 0.02)\; \text{theta band: } p = 0.06) emerged across all trials: Participants with higher compared to lower reasoning ability had increased alpha activity at frontal and fronto-medial channels. Furthermore, the larger alpha decrease in incongruent than congruent trials was largely driven by high reasoning ability participants \((p = 0.03)\).

The EEG results for the visual arrays task are depicted in Figure 3.7. Trials with load 4 compared to 2 were accompanied by higher alpha activity at left and right central to parietal channels \((p < 0.001; \text{theta band: } p = 0.06)\). Trials with load 6 compared to
3.3 Results

4 did not differ in either frequency band ($p > 0.10$). Furthermore, reasoning ability was positively related to theta activity at left-frontal ($p = 0.01$) and right parieto-occipital channels ($p = 0.04$) over all trials. In the alpha band, reasoning ability was associated with generally increased alpha activity ($p < 0.01$). In addition, the increase in alpha activity for load 4 compared to 2 seemed to be mainly due to participants with high reasoning ability, at least in the right hemisphere ($p < 0.001$). We did not find any

![Figure 3.6: Significant results of the cluster-based permutation analyses for the Stroop task. Statistically significant channels ($p < 0.05$) are indicated by asterisks (*). Tilde (~) signifies "predicted by"; INC = incongruent trials; CON = congruent trials; Diff. = difference; RA = Reasoning Ability.](image_url)

55
3.3 Results

Figure 3.7: Significant results of the cluster-based permutation analyses for the visual arrays task. Statistically significant channels ($p < 0.05$) are indicated by asterisks (*). Tilde (~) signifies "predicted by"; Diff. = difference; RA = Reasoning Ability.
Figure 3.8: Significant results of the cluster-based permutation analyses for the secondary memory task. Statistically significant channels \((p < 0.05)\) are indicated by asterisks (*). Tilde (\(~\)) signifies "predicted by"; MNT = maintenance; REC = recall; Diff. = difference; RA = Reasoning Ability.
3.3 Results

Figure 3.9: Significant results of the cluster-based permutation analyses for the updating task. Statistically significant channels (\(p < 0.05\)) are indicated by asterisks (*). Tilde (~) signifies "predicted by"; MNT = maintenance; UPD = updating; Diff. = difference; RA = Reasoning Ability.
3.4 Discussion

difference between load 6 and 4 with respect to reasoning ability.

The results for the secondary memory task are depicted in Figure 3.8. We found that load 3 compared to load 2 was associated with larger theta activity at fronto-medial (\( p < 0.01 \)) and further left-hemispheric channels (F7, FT7, T7, TP7, P7, CP5; \( p = 0.02 \)). This was accompanied by increased alpha activity at left central-parietal channels (\( p < 0.01 \)). Load 4 compared to load 3 was characterized by lower alpha activity at right-frontal and left-parietal channels (\( p < 0.001 \); theta band: \( p = 0.08 \)). During maintenance, participants with higher reasoning ability had more alpha activity at fronto-medial and parietal channels in both hemispheres across all loads (\( p < 0.01 \); theta band: \( p = 0.06 \)). Reasoning ability was also associated with increased alpha activity in trials with load 3 compared to load 2 at fronto-medial (\( p = 0.02 \)) and central-parietal channels in the right hemisphere (\( p = 0.02 \)). In contrast, reasoning ability was related to larger alpha activity decreases from load 3 to load 4 at left parietal (\( p = 0.02 \)) and right central channels (\( p = 0.03 \)). Finally, with respect to brain activity during recall, reasoning ability was associated with widespread increased alpha activity, in particular at fronto-medial and right-parietal channels (\( p < 0.01 \)).

Figure 3.9 shows the results for the updating task. Brain activity in the updating phase compared to the encoding phase was increased at fronto-medial and left-parietal channels in the theta band (\( p < 0.01 \)) and decreased at left parietal channels in the alpha band (\( p = 0.01 \)). In the updating phase, reasoning ability was related with higher theta activity at fronto-medial channels (\( p = 0.04 \)) and generally higher alpha activity (\( p < 0.01 \)). In the maintenance phase, participants with higher compared to lower reasoning ability were characterized by widespread increased alpha activity (\( p < 0.01 \)). Finally, reasoning ability was not associated with differences in brain activity during updating compared to maintenance.

3.4 Discussion

Interindividual differences in WMC are strongly related to fluid intelligence, but there is still debate about the cognitive mechanisms underlying this association. The most frequently discussed hypotheses stress the importance of individual differences in attentional control, primary memory, or retrieval from secondary memory. These mechanisms, however, have mainly been compared using performance measures alone. In
the present study, we complemented behavioral with neural data to shed light on these issues. We investigated the relationship between reasoning ability and oscillatory EEG activity in working memory tasks representative for the different theoretical accounts: 1) A color-word Stroop task as a measure of attention control, 2) a visual arrays task as a measure of primary memory, 3) a secondary memory task to measure retrieval from secondary memory, and 4) an updating task involving both attention control and primary memory processes. We focused on oscillatory activity in the theta (3-6 Hz) and alpha (8-12 Hz) frequency band due to their role in both intelligence and working memory (Neubauer & Fink, 2009; Sauseng et al., 2010). This approach provides a unique opportunity to gain a more fine-grained understanding of reasoning ability.

The behavioral data overall confirmed our expectations: We found that solution rates in the Stroop task were lower for incongruent compared to congruent trials (Hanslmayr et al., 2008) and decreased with increasing memory load in the visual arrays task (Luck & Vogel, 1997). The load-related effects in the secondary memory task, although not significant, were also pointing in the right direction: Later items were associated with lower performance than earlier items (Jensen et al., 2002). Furthermore, participants with higher reasoning ability generally performed better than participants with lower reasoning ability across all tasks. For the present study, however, the EEG data were more important and enlightening than the behavioral data.

The findings for the theta band are well in line with our hypotheses and its suggested role in control mechanisms in working memory (Bastiaansen & Hagoort, 2003; Klimesch, 1999; Roux & Uhlhaas, 2014; Sauseng et al., 2010). In the Stroop task, theta activity was increased in incongruent compared to congruent trials, in particular at fronto-medial channels, arguably reflecting inhibitory processes to suppress the dominant behavior (i.e., reading the word) in incongruent trials (Hanslmayr et al., 2008). Theta activity was also higher for load 3 compared to load 2 at fronto-medial channels in the secondary memory task (Gevins et al., 1997; Jensen et al., 2002; Sauseng et al., 2007). However, no difference was found for loads 4 and 3, possibly indicating that theta activity could not be further increased due to the close proximity to the supposed capacity limit. Furthermore, fronto-medial and parietal areas showed increased theta activity during updating compared to maintenance in updating task (Griesmayr et al., 2010). In the visual arrays task, in contrast to our expectations, the results did not show any difference in the theta band with respect to load (cf. Palva et al., 2010). This might
be explained by the fact that theta activity was actually moderated by reasoning ability, with participants with higher compared to lower reasoning ability showing more activity across all loads. Finally, in the updating task, reasoning ability was also associated with increased theta power during updating. Although not hypothesized, these interindividual difference effects in the theta band suggest that participants with higher reasoning ability generally activated more processes to coordinate cognitive resources in the visual arrays and updating tasks (Mitchell et al., 2008; Sauseng et al., 2007; Sauseng et al., 2010).

Interestingly, underlining the close relationship between alpha activity and intelligence, all other effects with respect to reasoning ability were restricted to the alpha band (Neubauer & Fink, 2009). Alpha oscillations are generally thought to reflect functional inhibition of brain areas, with lower activity indicating engagement and higher activity indicating disengagement of specific brain areas (Jensen & Mazaheri, 2010; Klimesch et al., 2007; Pfurtscheller & Lopes da Silva, 1999). We found that reasoning ability was associated with generally increased alpha activity in all tasks, confirming our expectations and mirroring a wealth of studies on the neural efficiency hypothesis of intelligence. Interpreted in the light of alpha activity reflecting functional inhibition (Jensen & Mazaheri, 2010; Klimesch et al., 2007), this fits with the assumption that reasoning ability is related to the disuse of task-irrelevant areas and a more focused use of task-relevant areas (Haier et al., 1988; Neubauer & Fink, 2009). The topographies generally suggest widespread cortical activation differences between participants with higher and lower reasoning ability, including frontal, central, parietal and occipital areas (Jung & Haier, 2007). The exact pattern of brain activation, however, depended on the task and the involved cognitive functions.

In addition to these general differences, we also examined whether brain activity differs between participants with higher and lower reasoning ability in the cognitive mechanisms suggested to be involved in WMC. In line with our expectations, there were differences with respect to multiple mechanisms. In the Stroop task, we found that the more demanding incongruent trials compared to congruent trials were accompanied by less alpha activity (e.g., Gevins et al., 1997). Crucially, however, this was mainly driven by participants with higher reasoning ability, suggesting that they are better able to activate brain resources to meet the increased attention control demand to suppress reading the word. In accordance with previous research, a load-dependent increase in
alpha activity from load 2 to 4 in the visual arrays task and from load 2 to 3 in the secondary memory task emerged (Jensen et al., 2002; Palva et al., 2010; Palva et al., 2011; Scheeringa et al., 2009). The present study, however, extends these findings by showing that the respective pattern of brain activity is more apparent in higher reasoning ability participants. This might reflect that reasoning ability is related to better disengagement of task-irrelevant areas, in this case central and occipital areas (Jensen & Mazaheri, 2010, but see also Palva et al., 2011). We also found no difference between loads 6 and 4 in the visual arrays task, which supports research showing that average brain activity reaches a plateau around load 4, arguably reflecting the capacity limit that is reached at that point (Sauseng et al., 2009; Todd & Marois, 2004; Vogel & Machizawa, 2004). However, in contrast to our predictions, alpha activity was actually lower in load 4 compared to load 3 in the secondary memory task. That this was also stronger in participants with higher reasoning ability suggests that they invested more brain resources, which at first glance is not in line with the neural efficiency hypothesis. However, Neubauer and Fink (2009) argued that task difficulty is an important moderator with respect to neural efficiency. Specifically, more intelligent individuals were speculated to "save" resources during tasks of easy and medium difficulty (thus confirming the hypothesis), but are able to invest more resources in tasks of higher (subjective) difficulty (see also Dunst et al., 2014; Nussbaumer et al., 2015). Thus, in the secondary memory task, more intelligent participants might have flexibly adapted their brain activity to the task demands by investing less for lower loads and more for higher loads.

Finally, we were interested in the reasons why updating tasks correlate so well with other working memory tasks and reasoning ability. Behavioral studies indicate that this might be more due to general maintenance and processing demands rather than the requirement to update the contents of working memory, i.e., the attention control component of the tasks (Ecker et al., 2010; Wilhelm et al., 2013). Reminiscent of these findings, the neural data of the present study also support the notion that updating per se is not the main cause: Although alpha activity was generally decreased in the more demanding updating phase compared to the maintenance phase (see e.g. Gevins et al., 1997), the magnitude of the change was unrelated to reasoning ability.

To sum up, the main interest of the present study was to illuminate the neurocognitive basis of WMC and reasoning ability. In accordance with recent studies, we expected that not only a single mechanism but multiple mechanisms are involved (Conway et al.,
3.4 Discussion

2011; Conway & Kovacs, 2015; Cowan et al., 2006; Shipstead et al., 2014; Unsworth & Engle, 2007; Unsworth & Spillers, 2010). The present findings, which showed that reasoning ability was related to alpha activity during tasks involving attention control, primary memory and retrieval from secondary memory, are in line with this. Although a differential functioning of multiple mechanisms is one plausible explanation, it could still be argued that a single common mechanism is responsible for these effects (e.g., a better management of attentional resources in general). However, further support for the multi-mechanism view comes from the findings that reasoning ability was also related to how alpha activity was adapted to the conditions within the tasks. For example, participants with higher reasoning ability were able to flexibly adapt their brain activity to deal with the increased demands on attention control when solving incongruent trials in the Stroop task and on primary memory when memory load increased in the visual arrays or secondary memory task.

Nonetheless, a shortcoming of the present study is that we were not able to investigate recall effects independent of already existing differences during maintenance. Future studies might try to ensure that all participants have enough time to perfectly maintain the information before investigating whether any recall-related effects occur. Another limitation is that we did not explore the relevant mechanisms within a single task but through several tasks. Although this approach allowed us to better link our findings to prior research, future studies should try to manipulate specific mechanisms (e.g., attention control) within a single task to have better control over the involved mechanisms (e.g., Gray, Chabris, & Braver, 2003; Vogel, McCollough, & Machizawa, 2005). Ideally, functional magnetic resonance imaging (fMRI) should be used in conjunction with EEG to better pinpoint any activation differences. This could potentially allow for the localization of individual differences in attention control, primary memory and secondary memory retrieval to specific brain networks (e.g., Conway et al., 2011). Finally, we want to stress that we do not argue to explain reasoning ability in its entirety. We are talking about the shared mechanisms between WMC and reasoning ability, which does not preclude that reasoning ability involves processes that go over and above those required in WMC tasks (see e.g., Harrison, Shipstead, & Engle, 2015; Shipstead, Harrison, & Engle, 2016).

The association between WMC and reasoning ability is assumed to arise from interindividual differences in attention control, primary memory, or retrieval from sec-
ondary memory. We aimed to better understand this association and investigated brain activity related to reasoning ability during working memory tasks involving these mechanisms. The results showed that brain activity, in particular in the alpha band, could differentiate between participants with higher and lower reasoning ability during all suggested cognitive mechanisms. This confirms recent research showing that the neuropsychological basis of WMC and reasoning ability is dependent on multiple mechanisms (e.g., Conway et al., 2011; Conway & Kovacs, 2015; Shipstead et al., 2014; Unsworth & Spillers, 2010). More generally, the present study also highlights the added value and potential of neural measures to test and compare psychological theories.

3.5 Acknowledgments

We would like to thank Caroline Wölfle for her assistance in collecting the EEG data.

References

References


References


68
References


References


References


References

research findings. Psychological Bulletin, 124(2), 262–274. doi:10.1037/0033-2909.124.2.262


References


4 The Relation Between Working Memory and Conceptual Learning: An EEG Study

Bruno Rütsche, Sarah Hofer, Elsbeth Stern

ETH Zurich, Research on Learning and Instruction

Abstract

In the present study, we were interested in the role of working memory as a resource for science learning in adolescents, since much of the previous research focused on mathematics learning in children. We thus investigated the relationship between a theoretically driven selection of working memory tasks and conceptual learning, specifically in physics, in a group of adolescents attending Swiss higher secondary schools. On top of that, we measured brain activity with EEG during working memory performance to examine whether neural measures increase our understanding of learning over and above behavioral measures. Statistical analysis revealed a unique relation between the performance in a secondary memory task, which required the retrieval of former information from long-term memory, and conceptual learning. The ability for controlled retrieval might have helped to retrieve the relevant prior knowledge and in turn to construct a coherent network of conceptual knowledge. In contrast, the ability to stay on task and inhibit irrelevant information as well as the maintenance and updating of information were not significantly related to learning. With respect to the neural data, we found no association with learning for either oscillatory power or functional network characteristics in the theta and alpha frequency bands over and above behavioral measures. Overall, these findings highlight the relevance of controlled retrieval for conceptual learning in an educational context. Further studies are required to better flesh out opportunities and limits of neural measures to inform learning in an educational context.
4.1 Introduction

*Keywords:* conceptual learning; physics; working memory capacity; controlled retrieval; secondary memory; electroencephalography; neuroimaging

*Author contributions:* BR was involved in the study design, acquisition of data, interpretation of data, and writing of the first draft of the manuscript. SH contributed to the acquisition of the data and critically revised the manuscript. ES contributed to the study design and the critical revision of the manuscript.

4.1 Introduction

Since the beginning of modern psychology, it has been a central endeavor in psychological research to quantify and explain interindividual differences in learning. Research generally suggests that working memory is an important determinant of academic achievement and knowledge acquisition (e.g., Alloway & Alloway, 2010; Gathercole, Pickering, Knight, & Stegmann, 2004). However, most of these investigations have been concerned with mathematics learning and language competencies in children (Peng, Namkung, Barnes, & Sun, 2016; Raghubar, Barnes, & Hecht, 2010). In the present study, we therefore investigated the relation between different working memory measures and physics learning in a group of adolescents attending Swiss higher secondary schools. We also measured brain activity with electroencephalography (EEG) during working memory performance, which provided an additional perspective with the potential to uniquely inform the understanding of this association.

4.1.1 The Role of Working Memory for Learning

Working memory is generally conceptualized as a capacity-limited system that allows for the retention of a small amount of information in an accessible state for the execution of cognitive tasks (Baddeley, 2012; Baddeley & Hitch, 1974; Cowan, 1999). According to the embedded-processes model (Cowan, 1988, 1999), working memory is an attentional system that directly acts upon long-term memory representations by increasing or decreasing their activation states. It is thought to consist of central executive processes and long-term memory representations differing with respect to their activation level. The most activated representations reside in the focus of attention (or primary
4.1 Introduction

memory), which is thought to be constrained with respect to the number of elements that can be maintained simultaneously. The less activated representations, in turn, are assumed to be located outside of the focus of attention in the activated portion of long-term memory (or secondary memory; see Unsworth & Engle, 2007). The importance of working memory for learning, simply defined as building structured and stable memory representations in long-term memory, is thus quite obvious. This holds especially true for complex material that relies on reasoning processes to integrate and abstract pieces of knowledge (e.g., Ackerman, 2007; Cowan, 2014; Primi, Ferrão, & Almeida, 2010; Stern, 2015; Voelkle, Wittmann, & Ackerman, 2006).

One area where this clearly applies is the acquisition of conceptual knowledge in STEM (Science, Technology, Engineering, Mathematics) fields. Conceptual knowledge is generally thought to consist of abstract and relational representations of the core principles of a particular domain and their interrelations, flexibly organized in a hierarchical fashion (e.g., Schneider & Stern, 2010). For example, an abstract understanding of force in Newton’s second law \( F = m \cdot a \) requires an integration of the concepts of velocity \( v = \frac{s}{t} \), acceleration \( a = \frac{v}{t} \), and the multiplicative relation of mass and acceleration. This abstraction makes it possible that the knowledge is not tied to a specific context, i.e., the context where it was acquired, but can be flexibly applied in many different situations (Chi & VanLehn, 2012). Although acquiring such transferable knowledge is one of the main goals of education, it is difficult and many students fail to do so (Clement, 1982; Halloun & Hestenes, 1985; Hestenes, Wells, & Swackhamer, 1992). In this study, we investigated the general role of working memory as a resource for conceptual learning.

Based on theoretical considerations about the nature of the capacity limit of working memory, there are several means through which working memory could influence conceptual learning (Cowan, 2014). The attention control hypothesis stresses the fact that individuals with higher working memory capacity (WMC) are better at staying on task and controlling their attention to maintain or suppress information, particularly in the face of interference or distraction (Engle, 2002; Engle, Tuholski, Laughlin, & Conway, 1999). For example, this circumstance is thought to become noticeable in experimental settings where a habitual response has to be inhibited such as in the classic Stroop task. The goal in this task is to name the ink color in which color words are written (e.g., the word "RED" printed in red or green ink). In cases where the ink color and color word
4.1 Introduction

do not match, the tendency to answer based on reading the word has to be overcome to arrive at the correct solution, resulting in generally worse performance. Kane and Engle (2003) showed that WMC particularly plays a role in contexts where the likelihood to temporarily forget the overall task goal (i.e., name the ink color) is increased by including less trials with a mismatch between written word and ink color. In these situations, participants with higher compared to lower WMC were especially successful, arguably because they are better able to control their attention and stay on task. Further evidence for the importance of attention control comes from a more recent study by Mrazek et al. (2012). They found that, using an experience-sampling approach, participants who indicated better maintenance of on-task thoughts and less mind-wandering during laboratory testing performed better on measures of WMC, intelligence, and school achievement.

In the current study, we used a color-word Stroop task as a measure of attention control.

A different strand of theories attribute the limit of WMC to restrictions in storage capacity, i.e., the maximum number of elements or relations that can be held in primary memory (Cowan, 2001; Cowan et al., 2005; Halford, Cowan, & Andrews, 2007; Halford, Wilson, & Phillips, 1998; Luck & Vogel, 1997). With respect to learning in educational contexts, the relational complexity theory specifically suggests that concept learning is limited by the number of variables that can be represented and processed in parallel (Halford et al., 1998). For example, in order to learn and understand the concept of velocity, distance and time as well as their relation have to be simultaneously represented in working memory. The concept "velocity" can then be flexibly used to build the higher-level concept "acceleration" (i.e., velocity changes over time) without overloading working memory (Halford et al., 1998). In the present study, we used the visual arrays task as a measure of storage capacity (based on Cowan et al., 2005). The task required the maintenance of up to 6 colored squares for a short time period.

Another often used type of working memory task requires that a relatively small number of items, typically below the proposed capacity limit of primary memory, are maintained and repeatedly updated. These so-called updating tasks have generally been found to be better predictors of learning than other executive function tasks, at least in children (Bull & Lee, 2014; van der Ven, Kroesbergen, Boom, & Leseman, 2012). However, the reason for this is still a matter of debate. Updating tasks have been used as both indicators of executive function (Miyake & Friedman, 2012; Miyake et al., 2000) and WMC, as they also require the maintenance of information for successful perfor-
4.1 Introduction

mance (Kyllonen & Christal, 1990; Wilhelm, Hildebrandt, & Oberauer, 2013). Recent research suggests that the latter seems to be the main reason why updating tasks correlate well with both other working memory and reasoning tasks (Ecker, Lewandowsky, Oberauer, & Chee, 2010; Wilhelm et al., 2013). This is well explained by the working memory model by Oberauer (2009). In this model, WMC does not arise from limits in the number of elements per se but from interference between bindings (e.g., binding of items to positions in working memory tasks or concepts to roles in propositional schemata). Thus, updating tasks tax WMC since they require the continuous and flexible maintenance of temporary bindings. Regardless of the theoretical explanation, the theories either emphasizing storage capacity or binding limits agree that WMC restricts the complexity of structural representations, which in turn constrains conceptual learning. In the present study, we therefore also included an updating task. The task required that 3 positions in a 4×4 grid had to be continuously maintained and updated (based on arrows pointing either left, right, up or down).

A further factor that was suggested to affect WMC is the controlled retrieval of information (Unsworth & Engle, 2007). Once the limit of primary memory has been reached, the information is thought to be displaced to secondary memory, from where it can be reinstated by a probabilistic cue-dependent search process. Participants with higher WMC are argued to apply more selective retrieval cues, which allows them to recall the relevant information with higher probability. For example, in cued recall, Unsworth (2009) found that they overall recalled more items and suffered less from intrusions than participants with lower WMC. In the present study, we used a secondary memory task to measure the ability for controlled retrieval. The task required to maintain 4 positions that were sequentially highlighted within a 4×4 grid and afterwards recall them in serial order. To ensure that the information was indeed recalled from secondary memory, an unrelated processing task had to be performed before recall.

Overall, research with children suggests that working memory is moderately related to learning and achievement in mathematics (Peng et al., 2016). The association seems to be stronger for updating compared to inhibition tasks (e.g., Stroop task; Bull & Lee, 2014). The few studies that investigated the association of working memory and academic achievement in adults also suggest a link of moderate strength (Cowan et al., 2005; Rohde & Thompson, 2007; Tolar, Lederberg, & Fletcher, 2009). However, most of this research has focused on mathematics, while the role of working memory for sci-
4.1 Introduction

ence learning, particularly in adolescents, is not well known. Furthermore, to the best of our knowledge, the contributions of the different working memory functions outlined above to conceptual learning have not been simultaneously assessed. The present study is unique in the sense that we investigated the association between working memory tasks involving these functions and conceptual learning in a group of adolescents. In particular, the relevance of the controlled retrieval from secondary memory has not yet been examined. This function seems especially crucial because the ability to retrieve the relevant prior knowledge from long-term memory is an indisputable prerequisite for successful conceptual learning, as evidenced by a wealth of findings showing that prior knowledge is the most important predictor of future learning (Ackerman, 2007; Carey, 2000; Stern, 2015; Tricot & Sweller, 2013).

4.1.2 The Added Value of Neuroimaging

In the last decades, the development of increasingly sophisticated methods to measure both brain structure and function has permitted invaluable insights into how the brain works. However, this development has also been accompanied by an extensive debate about the relationship between education and neuroscience (e.g., Ansari & Coch, 2006; Bowers, 2016; Bruer, 1997; De Smedt, Grabner, Kadosh, & Dowker, 2015; Gabrieli, 2016; Gabrieli, Ghosh, & Whitfield-Gabrieli, 2015; Goswami, 2006; Howard-Jones et al., 2016; Stern, 2005; Stern & Schneider, 2010; Stern, Grabner, & Schumacher, 2016; Szücs & Goswami, 2007). While some deny any added value of neuroscientific research for education (for a recent critique see e.g. Bowers, 2016), most researchers share a moderate view. They agree that, while the merit of neuroscientific findings for educational practice is currently clearly limited, the application of neuroimaging can inform educational research in some cases.

One promising approach that has gained popularity in recent years is the use of neuroscientific measures to predict individual differences in future performance or the response to educational interventions (Ansari & Lyons, 2016; De Smedt et al., 2015; Gabrieli et al., 2015). For example, Hoeft et al. (2011) compared behavioral and neural indices of developmental dyslexia and found that only the latter were associated with future reading gains. Dumontheil and Klingberg (2012) reported that brain activity measured with functional magnetic resonance imaging (fMRI) during working memory
4.1 Introduction

tasks predicted arithmetic performance two years later even after reasoning and working memory performance were controlled for. Another study showed that both brain structure and functional connectivity, but no behavioral measure such as intelligence, working memory and mathematical ability, were related to learning gains in an eight-week one-to-one math tutoring program in children (Supekar et al., 2013). Similarly, Evans et al. (2015) showed that the growth in numerical abilities in children over six years is better predicted by brain structure and functional connectivity than behavioral measures.

In the current study, we wanted to investigate in a similar fashion whether neural activity during working memory performance is correlated with conceptual learning in an educational context over and above behavioral measures. The application of neuroimaging has the potential to increase our understanding of the relation between specific working memory functions and conceptual learning. Therefore, we analyzed EEG data recorded during working memory tasks with respect to both power and functional connectivity. EEG power, on the one hand, reflects the amount of synchronously discharging neurons in local neuronal networks (i.e., power increases indicate increases in synchronous activity). However, not only local processing but also the functional integration of information processed in separate brain areas is crucial for cognitive processing (Fries, 2005; Varela, Lachaux, Rodriguez, & Martinerie, 2001). Functional connectivity in electroencephalographic data is assumed to be established when the phases of signals from different areas are synchronized, i.e., when their phases are constant. We used the debiased weighted phase lag index (wPLI), a recently developed measure unaffected by volume conduction and common noise sources, to characterize phase synchronization (Vinck, Oostenveld, van Wingerden, Battaglia, & Pennartz, 2011). We then applied graph theory to quantify the structure of the functional network, using the mean clustering coefficient as a measure of functional segregation and the characteristic path length as a measure of functional integration (Rubinov & Sporns, 2010). We focused on the theta and alpha bands due to their essential role for (working) memory processes (Grabner, Fink, Stipacek, Neuper, & Neubauer, 2004; Jensen, Gelfand, Kounios, & Lisman, 2002; Klimesch, 1999; Klimesch, Freunberger, Sauseng, & Gruber, 2008; Osipova et al., 2006; Roux & Uhlhaas, 2014; Sauseng, Griesmayr, Freunberger, & Klimesch, 2010). Increased theta activity in working memory is generally assumed to serve the control and coordination of cognitive resources (Roux & Uhlhaas, 2014;
4.1 Introduction

Sauseng, Hoppe, Klimesch, Gerloff, & Hummel, 2007; Sauseng et al., 2010), while increased alpha activity is assumed to reflect local cortical inhibition (Jensen & Mazaheri, 2010; Klimesch, Sauseng, & Hanslmayr, 2007). For example, more intelligent subjects are often characterized by increased alpha activity, which is interpreted to reflect that intelligence is related to neural efficiency (Neubauer & Fink, 2009). Both the theta and alpha bands have also been previously suggested as relevant neural markers to assess instructional interventions (Antonenko, Paas, Grabner, & Gog, 2010).

4.1.3 The Present Study

The first goal of the present study was to investigate the relationship between a theoretically driven selection of working memory tasks and conceptual learning, specifically in physics, in a group of adolescents. A color-word Stroop task was used to measure staying on task and inhibition of irrelevant information, a visual arrays and updating task were used to measure the maintenance and updating of information, and a secondary memory task was used to measure the controlled retrieval from secondary memory. To have an ecologically valid measure of learning, a subsample of participants was recruited from a study that was performed in classrooms in Swiss higher secondary schools (Hofer, Schuhmacher, Rubin, & Stern, 2017). This study collected conceptual knowledge about Newtonian mechanics with the basic Mechanics Conceptual Understanding (bMCU) test prior to and after an 18-lesson instruction (Hofer, Schumacher, & Rubin, 2017). Generally, we expected the performance in all working memory tasks to be positively related to conceptual learning. However, the secondary memory task, due to its explicit link between working and long-term memory via the requirement for controlled retrieval, was hypothesized to be uniquely related to conceptual learning. The ability to retrieve relevant knowledge is clearly a crucial step for gradually developing an abstract understanding of the involved concepts and their interrelations.

The second goal of this study was concerned with the added value of neural measures to inform our understanding of conceptual learning over and above behavioral measures. Therefore, we also recorded EEG during working memory performance and investigated the association between brain activity and conceptual learning in an exploratory fashion. We analyzed EEG data with respect to both power and functional connectivity in the theta and alpha frequency bands to quantify both local and distributed
4.2 Method

4.2.1 Participants

All participants attended Swiss higher secondary schools (Gymnasium)\(^1\) and were recruited from a classroom study conducted in our research group (Hofer, Schuhmacher, et al., 2017). This previous study aimed at investigating the increase in conceptual knowledge about Newtonian mechanics as a consequence of an instruction encompassing 18 lessons. Specifically, half of the participants received an instruction that applied selected methods proven to stimulate the construction of conceptual knowledge such as generating and inventing own solutions (Kapur, 2014; Schwartz & Martin, 2004), comparing and contrasting cases (Gick & Holyoak, 1983; Schalk, Saalbach, & Stern, 2016; Ziegler & Stern, 2014), giving self-explanations (Chi, Leeuw, Chiu, & LaVancher, 1994), and asking meta-cognitive questions (Beeth, 1998), while the other half received conventional instruction about the same topic. We focused on the individual learning gains irrespective of group membership in the present study, although the former group slightly outperformed the latter group with respect to conceptual knowledge.

From the classroom study, we adopted the pre- (before instruction) and posttest (after instruction) scores in the bMCU test (Hofer, Schumacher, & Rubin, 2017) as a measure of conceptual learning and the Raven’s Advanced Progressive Matrices (RAPM; Raven, Raven, & Cout, 2003) as a measure of reasoning ability. However, due to missing values present in either of those tests, 9 participants were excluded from the initial subsample of 39, resulting in a final sample of 30 participants (21 females; \(M_{\text{age}} = 16, SD_{\text{age}} = 0.91\)). The same sample was used in a related study (Rütsche, Hofer, & Stern, 2017), but there the focus was on the relation between WMC and reasoning ability instead of conceptual learning. The university ethics review board approved the study and written informed consent was provided by all participants’ parents. The participation

---

\(^{1}\)Swiss Gymnasium is a higher secondary school type that corresponds to US college preparatory classes and is attended by only approximately 20% of Swiss students.
was reimbursed with 100 CHF (about 100 USD) plus travel expenses.

4.2 Method

4.2.2 Procedure

The participants were informed about the present investigation and had the opportunity to sign up for it at the end of the classroom study. Individual sessions, which lasted about 3.5 hours, were then scheduled and conducted in the EEG laboratory. After mounting the EEG system, the participants were seated in an electromagnetically shielded cabin and the recording quality was checked and optimized. All participants then performed the same order of tasks while EEG was recorded: operation span, updating, visual arrays, secondary memory, and Stroop task. Self-timed pauses were allowed both between as well as within tasks. The EEG system was then removed and the participants were asked to solve several subscales from the I-S-T 2000 R intelligence test (Liepmann, Beauducel, Brocke, & Amthauer, 2007). These scales, as well as data from the operation span task (adapted from Lewandowsky, Oberauer, Yang, & Ecker, 2010), which had the purpose to familiarize the participants with the experimental setting, were not used in the present study. Finally, we thanked all subjects for taking part and dismissed them after a thorough debriefing.

4.2.3 Measures

**basic Mechanics Conceptual Understanding (bMCU)**

The bMCU test is a Rasch-conform multi-choice test to measure conceptual understanding of Newtonian mechanics (Hofer, Schumacher, & Rubin, 2017). The items cover the topics *inertia and motion*, *force and acceleration*, *balance of forces*, and *reciprocal action* and are scored as correct only if all right but none of the wrong response alternatives are selected. Figure 4.1 depicts a sample item. The short version with 11 items was used at the pretest, while the long version with 17 items was used at the posttest.

**Raven’s Advanced Progressive Matrices test (RAPM)**

The RAPM (Raven et al., 2003) is a nonverbal test to measure abstract reasoning ability and fluid intelligence. The test consists of 48 items and each item requires that a pattern within a 3×3 matrix is completed by identifying the missing element. The 12 items
4.2 Method

Figure 4.1: Sample item of the basic Mechanics Conceptual Understanding (bMCU). Answers 3 and 4 are correct.

forming Set I were used to accustom the participants to the task, while the 36 items forming Set II were administered as test items with a standard time constraint of 40 min. The score across all Set II items was used for statistical analysis.

Working Memory Tasks

Stroop task: The task consisted of 270 trials in which the German color words ROT (RED), GRÜN (GREEN), or BLAU (BLUE) were shown in either red, green, or blue ink (see Figure 4.2). The word and ink color were matching in 162 trials (congruent trials; 60%), while they were not matching in 108 trials (incongruent trials; 40%). The participants were instructed to answer based on the ink color (while ignoring the color word) and press the correspondingly colored button on a response box (Psychology Software Tools, Sharpsburg, PA). The trial then ended with the presentation of a fixation point for a random duration between 1.25 and 1.75s.

Visual arrays task: Each trial in the visual arrays task began with the brief presentation (0.1s) of an array containing 2, 4, or 6 colored squares (adapted from Cowan et al., 2005; Vogel & Machizawa, 2004, see Figure 4.2). After a delay period (0.9s), a
4.2 Method

second array was shown that was either identical to the first one or differed in the color of a single square. The potential change was marked by a circle surrounding the square. The task was to indicate for each trial whether the two arrays were the same or not via two labeled buttons. A fixation point was shown after each trial for a random duration between 1.25 and 1.75s. Overall, 240 trials were presented, consisting of an equal number of trials from each load (i.e., 2, 4, and 6 squares). The color changed from the first to the second array in half of the trials.

Secondary memory task: In the secondary memory task, each trial required that the positions of 4 sequentially presented squares within a 4×4 grid had to be maintained (presentation duration: 0.5s, interstimulus interval: 1s; see Figure 4.2). After that, 4 simple arithmetic equations (adopted from Lewandowsky et al., 2010) were shown which had to be judged to be either correct (right button) or wrong (left button). The participants were then asked to recall all maintained square positions and press a button, upon which they selected the positions (in serial order) by clicking in the respective cells. Overall, 35 trials were presented, which were separated by a fixation point for a random duration between 2.75 and 3.25s.

Updating task: Each trial in the updating task started with a red, a green, and a blue square that were shown one after the other in a 4×4 grid (presentation duration: 0.5s, interstimulus interval: 1s; see Figure 4.2). This was followed by 4, 5, or 6 colored arrows that were presented in succession at the center of the grid. The color of the arrow singled out the relevant square for this updating step, while its orientation (either left, right, up or down) specified the direction the selected square had to be mentally moved. After all updating steps, the final position of each color had to be selected by clicking in the respective cells of the grid. 24 trials were presented overall, with one third of the trials either involving 4, 5, or 6 updating steps. A fixation point with a random duration between 2.75 and 3.25s was shown between the trials.

PsychoPy v1.81.00 (Peirce, 2007, 2009) was used to program and present all working memory tasks. Each task began with a series of training trials that included performance feedback. The stimuli for each task, i.e., the combination of colors, locations, and/or arrow directions, were randomly generated, but all participants received the same random sequence of stimuli. A more detailed description of the tasks can be found in Rütsche et al. (2017).
4.2 Method

Figure 4.2: Schematic depiction of the working memory tasks.

4.2.4 Electroencephalography (EEG)

We used a BioSemi ActiveTwo system (BioSemi, Amsterdam, The Netherlands) with 64 active electrodes to collect the EEG data. The feedback loop between the Common Mode Sense active and Driven Right Leg passive electrode formed the recording reference. The direct current offsets were kept below 40 mV (if possible below 20 mV) during the EEG recording.

EEG data were preprocessed in FieldTrip (version 02-10-2016; 64 bit MATLAB 2016a; Oostenveld, Fries, Maris, & Schoffelen, 2011). After importing, the data was filtered with a Butterworth high-pass filter at 0.5 Hz (filter order: 5, two passes) and a band-stop filter between 48 and 52 Hz (and its harmonic between 98 and 102 Hz) to remove DC drift and power line noise. The initial sampling rate of 1024 Hz was reduced to 256 Hz to speed up further processing. The data were then scanned for noisy channels and segments with artifacts: Bad data segments were first marked with FieldTrips (semi-)automatic procedures, which were then reviewed and potentially extended by manual inspection. We ran an Independent Component Analysis (ICA) on the cleaned data.
4.2 Method

separately for each subject and task and removed components indicating eye blinks and vertical/horizontal eye movements (Delorme & Makeig, 2004). After interpolating the previously rejected bad channels using spherical splines, we computed the average reference and epoched the data.

Fourier analysis was applied to calculate the time-frequency representations of the individual epochs for the frequency range from 3 to 12 Hz. We used a single Hanning taper and an adaptive time window of 3 cycles for each frequency ($\Delta T = \frac{3}{f}$), which slided across the epochs in steps of 50 ms (Osipova et al., 2006). We analyzed EEG data with respect to both power and functional connectivity. Specifically, we computed the wPLI, which is an advancement of the phase lag index (Stam, Nolte, & Daffertshofer, 2007), between all channel pairs to quantify phase synchronization (Vinck et al., 2011). The wPLI compared to PLI has been shown to be less sensitive to volume conduction, common noise sources (e.g., a common reference) and to have increased statistical power (Vinck et al., 2011).

Finally, we aggregated the power and wPLI values over epochs and time. As in our previous study that used the same sample (Rütsche et al., 2017), we aggregated across time from 0.4s to 0.7s after the appearance of the color word in the Stroop task (Hanslmayr et al., 2008) and from -0.6s to 0s before the appearance of the second array in the visual arrays task (Palva, Monto, Kulashekhar, & Palva, 2010; Vogel & Machizawa, 2004). In the updating task, we averaged samples from 0.2s to 1.5s after the presentation of the colored arrows and in the secondary memory task, we averaged across time from 0.2s to 1.5s for the individual maintenance stimuli and from 0.2s until 0.25s before the button press in the recall phase.

4.2.5 Statistical Analysis

Regarding the behavioral data, we were interested in the relation between the performance in the working memory tasks and conceptual learning. We therefore ran a multiple regression with the bMCU posttest score (bMCU post) as the criterion variable and the mean-centered average solution rates in the Stroop, visual arrays, updating, and secondary memory task as predictor variables. We additionally included three control variables: the mean-centered bMCU pretest score (bMCU pre), the mean-centered RAPM score in Set II as a measure of reasoning ability and the physics instruction con-
4.2 Method

dition (dummy coded: conventional instruction = 0, cognitively activating instruction = 1). However, to make sure that it was appropriate to average across all trial types for each task, we first ran a series of linear mixed effects regressions (Bates, Mächler, Bolker, & Walker, 2015). We investigated for each task whether the effect of bMCU post on performance, while controlling for bMCU pre, was dependent on the trial type (i.e., congruent vs. incongruent trials in the Stroop task; trials with 2, 4, or 6 squares in the visual arrays task; trials with 4, 5 or 6 updating steps in the updating task; 1st, 2nd, 3rd, or 4th item to maintain in the secondary memory task). Repeated measurements were accounted for by including a random intercept for each subject and statistical significance was assessed with "type 3" tests via the Kenward-Roger approximation (Højsgaard, 2016; Singmann, Bolker, Westfall, & Aust, 2016).

We used cluster-based permutation tests implemented in FieldTrip to perform statistical analysis of oscillatory power (Maris, 2012; Maris, Schoffelen, & Fries, 2007; Oostenveld et al., 2011). This approach has the advantage that the multiple comparison problem is alleviated because significance testing is based on a single overall test statistic (the so-called cluster-level statistic). In the present study, we first quantified an effect of interest with a $t$-statistic at each channel, which were then thresholded (at the 2.5th and the 97.5th quantiles), clustered based on spatial adjacency and summed within a cluster to obtain the cluster-level statistic. Statistical significance was computed by comparing the maximum cluster-level statistic with a reference distribution based on 3000 random draws using the Monte Carlo method.

We analyzed functional brain connectivity with graph theoretical measures implemented in the Brain Connectivity Toolbox (Bullmore & Sporns, 2009; Bullmore & Sporns, 2012; Rubinov & Sporns, 2010). We opted for a weighted network analysis, with channels representing nodes and wPLI values between two channels representing edges, instead of arbitrarily tresholding the connectivity matrix to yield binary networks (Stam et al., 2009). The mean clustering coefficient and the characteristic path length are two basic network parameters. The clustering coefficient measures local connectedness and reflects the average "intensity" of all triangles associated with each node (Rubinov & Sporns, 2010). The mean clustering coefficient (Cw) can then be regarded as a measure of overall functional segregation in the network. In contrast, the characteristic path length (Lw) is the average shortest path length between all pairs of nodes and reflects a measure of global integration in the network.
4.3 Results

To investigate the relation between neural activity and conceptual learning, we predicted bMCU post by either power (to obtain the channel-wise t-statistics; see above), mean clustering coefficient, or characteristic path length. Similar to the behavioral data, we also controlled for potential effects of bMCU pre, reasoning ability, instruction condition, and overall performance in the respective working memory task. Because the focus of the present study was on conceptual learning, we do not report the effects of the control variables.

Table 4.1: Descriptive statistics of the behavioral measures.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>bMCU post</td>
<td>4.43</td>
<td>2.01</td>
<td>1.00</td>
<td>9.00</td>
</tr>
<tr>
<td>bMCU pre</td>
<td>2.60</td>
<td>1.19</td>
<td>0.00</td>
<td>5.00</td>
</tr>
<tr>
<td>RAPM</td>
<td>27.63</td>
<td>4.82</td>
<td>19.00</td>
<td>34.00</td>
</tr>
<tr>
<td>Stroop</td>
<td>0.96</td>
<td>0.03</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>Visual arrays</td>
<td>0.90</td>
<td>0.04</td>
<td>0.82</td>
<td>0.98</td>
</tr>
<tr>
<td>Updating</td>
<td>0.80</td>
<td>0.16</td>
<td>0.25</td>
<td>1.00</td>
</tr>
<tr>
<td>Secondary memory</td>
<td>0.66</td>
<td>0.15</td>
<td>0.31</td>
<td>0.89</td>
</tr>
</tbody>
</table>

bCMU = basic Mechanics Conceptual Understanding; pre = pretest, post = posttest; RAPM = Raven’s Advanced Progressive Matrices.

4.3 Results

4.3.1 Performance

Table 4.1 gives the descriptive statistics and Table 4.2 the correlations of the behavioral measures. We first predicted the performance in each working memory task by bMCU post, trial type, and their interaction using linear mixed effects regressions, while controlling for the effect of bMCU pre. We found that conceptual learning was related to performance in the secondary memory task ($F(1, 27) = 4.82, p = 0.04$), but not the Stroop ($F(1, 27) = 1.03, p = 0.32$), visual arrays ($F(1, 27) = 2.95, p = 0.10$) or updating task ($F(1, 27) = 0.18, p = 0.67$). Trial type was significant in the Stroop ($F(1, 27) = 45.45, p < 0.01$), visual arrays ($F(1, 56) = 93.50, p < 0.01$) and updating task ($F(1, 56) = 3.77, p = 0.03$), while no effect was found in the secondary memory task ($F(1,$
4.3 Results

More importantly, the interaction was not significant in any task (Stroop: $F(1, 28) = 1.90, p = 0.18$; visual arrays: $F(2, 56) = 0.05, p = 0.95$; updating: $F(2, 56) = 2.64, p = 0.08$; secondary memory: $F(3, 84) = 1.21, p = 0.31$). Thus, we used the average performance in each task for all further statistical analyses. Furthermore, the control variable bMCU pre was related to the performance in the visual arrays ($F(1, 27) = 5.00, p = 0.03$), but not the Stroop ($F(1, 27) = 0.68, p = 0.42$), updating ($F(1, 27) = 0.72, p = 0.41$), or secondary memory task ($F(1, 27) = 0.92, p = 0.35$).

Table 4.2: Correlation matrix of all behavioral measures.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 bMCU post</td>
<td></td>
<td>0.56*</td>
<td>0.34</td>
<td>0.10</td>
<td>-0.10</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>2 bMCU pre</td>
<td></td>
<td></td>
<td>0.22</td>
<td>-0.07</td>
<td>-0.27</td>
<td>-0.14</td>
<td>0.06</td>
</tr>
<tr>
<td>3 RAPM</td>
<td></td>
<td></td>
<td></td>
<td>0.43*</td>
<td>0.44*</td>
<td>0.52*</td>
<td>0.40*</td>
</tr>
<tr>
<td>4 Stroop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.47*</td>
<td>0.38*</td>
<td>0.23</td>
</tr>
<tr>
<td>5 Visual arrays</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.51*</td>
<td>0.35</td>
</tr>
<tr>
<td>6 Updating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.58*</td>
</tr>
<tr>
<td>7 Secondary memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

bCMU = basic Mechanics Conceptual Understanding; pre = pretest, post = posttest; RAPM = Raven’s Advanced Progressive Matrices; *p < 0.05.

Therefore, to investigate the unique effect of each task on conceptual learning, we estimated the effects of the average performance in each task on bMCU post, while controlling for the effects of bMCU pre, RAPM, and instruction condition. We found that only performance in the secondary memory task uniquely predicted conceptual learning ($\beta = .48, 95\%-CI [0.05, 0.77], t = 2.25, p = 0.04$); all the other parameter estimates were not significant (see Table 4.3). To substantiate this finding, we compared this full model to a more restrictive model without secondary memory performance. We assessed model fit with the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), where relatively lower values indicate better model fit. Both criteria penalize the number of parameters included in a model to deal with the trade-off between goodness-of-fit and model complexity. The full model including secondary memory performance explained about 11% of unique variance (overall model: $32\% R^2$) and provided a better fit to the data than the more restrictive model (AIC: 121 vs. 134; BIC = 125 vs. 136).
4.3 Results

Table 4.3: Results of the regression model predicting bMCU post by working memory task performance.

<table>
<thead>
<tr>
<th>Working memory tasks</th>
<th>est</th>
<th>est se</th>
<th>lower</th>
<th>higher</th>
<th>est std</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroop</td>
<td>2.02</td>
<td>12.77</td>
<td>−24.47</td>
<td>28.50</td>
<td>0.03</td>
<td>0.16</td>
<td>0.88</td>
</tr>
<tr>
<td>Visual arrays</td>
<td>10.62</td>
<td>10.48</td>
<td>−11.11</td>
<td>32.36</td>
<td>0.20</td>
<td>1.01</td>
<td>0.32</td>
</tr>
<tr>
<td>Updating</td>
<td>−4.31</td>
<td>2.68</td>
<td>−9.88</td>
<td>1.26</td>
<td>−0.34</td>
<td>−1.61</td>
<td>0.12</td>
</tr>
<tr>
<td>Secondary memory</td>
<td>5.33</td>
<td>2.37</td>
<td>0.41</td>
<td>10.25</td>
<td>0.41</td>
<td>2.25</td>
<td>0.04</td>
</tr>
<tr>
<td>Control variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bMCU pre</td>
<td>0.88</td>
<td>0.29</td>
<td>0.28</td>
<td>1.47</td>
<td>0.52</td>
<td>3.07</td>
<td>0.01</td>
</tr>
<tr>
<td>RAPM</td>
<td>0.08</td>
<td>0.08</td>
<td>−0.10</td>
<td>0.25</td>
<td>0.18</td>
<td>0.93</td>
<td>0.36</td>
</tr>
<tr>
<td>Inst. Condition</td>
<td>0.76</td>
<td>0.68</td>
<td>−0.66</td>
<td>2.17</td>
<td>0.17</td>
<td>1.11</td>
<td>0.28</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>3.88</td>
<td>0.58</td>
<td>2.68</td>
<td>5.07</td>
<td>0.00</td>
<td>6.74</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

bCMU = basic Mechanics Conceptual Understanding; pre = pretest, post = posttest; RAPM = Raven’s Advanced Progressive Matrices; est = estimate; est se = estimate standard error, CI = confidence interval, est std = standardized estimate.

4.3.2 EEG Power / Connectivity

We applied cluster-based permutation tests to investigate the relation between conceptual learning and EEG power in the theta and alpha bands, with bMCU pre, RAPM, instruction condition, and overall performance in the respective task as control variables. The cluster algorithm did not identify any clusters in the Stroop, visual arrays, updating, and secondary memory tasks, hence no significance values can be provided.

Functional connectivity of the network was quantified with the mean clustering coefficient and the characteristic path length. We investigated the association between both network measures and bMCU post separately for each working memory task and frequency band (using the same control variables as for the cluster-based permutation analysis above). We again found no significant relation between EEG measures and conceptual learning (all \(p\)-values > 0.05). Table 4.4 shows the relevant parameter estimates.
Table 4.4: Parameter estimates of the relation between bMCU post and functional network parameters for the theta (3-6 Hz) and alpha bands (8-12 Hz).

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Parameter</th>
<th>est std</th>
<th>lower</th>
<th>higher</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>est 95% CI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta</td>
<td>Cw</td>
<td>−0.17</td>
<td>−0.50</td>
<td>0.20</td>
<td>−1.03</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>0.13</td>
<td>−0.20</td>
<td>0.50</td>
<td>0.76</td>
<td>0.45</td>
</tr>
<tr>
<td>alpha</td>
<td>Cw</td>
<td>−0.1</td>
<td>−0.40</td>
<td>0.20</td>
<td>−0.59</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>−0.08</td>
<td>−0.40</td>
<td>0.30</td>
<td>−0.43</td>
<td>0.67</td>
</tr>
<tr>
<td>Visual Arrays</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta</td>
<td>Cw</td>
<td>−0.32</td>
<td>−0.60</td>
<td>0.00</td>
<td>−1.96</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>0.26</td>
<td>−0.07</td>
<td>0.60</td>
<td>1.54</td>
<td>0.14</td>
</tr>
<tr>
<td>alpha</td>
<td>Cw</td>
<td>−0.11</td>
<td>−0.50</td>
<td>0.20</td>
<td>−0.63</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>0.19</td>
<td>−0.10</td>
<td>0.50</td>
<td>1.16</td>
<td>0.26</td>
</tr>
<tr>
<td>Updating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta</td>
<td>Cw</td>
<td>−0.31</td>
<td>−0.67</td>
<td>0.05</td>
<td>−1.66</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>0.1</td>
<td>−0.26</td>
<td>0.50</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>alpha</td>
<td>Cw</td>
<td>−0.21</td>
<td>−0.58</td>
<td>0.20</td>
<td>−1.09</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>0.34</td>
<td>−0.04</td>
<td>0.70</td>
<td>1.78</td>
<td>0.09</td>
</tr>
<tr>
<td>Secondary memory: encoding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta</td>
<td>Cw</td>
<td>−0.09</td>
<td>−0.43</td>
<td>0.30</td>
<td>−0.52</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>0.02</td>
<td>−0.32</td>
<td>0.40</td>
<td>0.11</td>
<td>0.91</td>
</tr>
<tr>
<td>alpha</td>
<td>Cw</td>
<td>−0.29</td>
<td>−0.62</td>
<td>0.04</td>
<td>−1.7</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>0.25</td>
<td>−0.06</td>
<td>0.60</td>
<td>1.57</td>
<td>0.13</td>
</tr>
<tr>
<td>Secondary memory: recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta</td>
<td>Cw</td>
<td>−0.15</td>
<td>−0.46</td>
<td>0.20</td>
<td>−0.92</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>0.09</td>
<td>−0.25</td>
<td>0.40</td>
<td>0.5</td>
<td>0.62</td>
</tr>
<tr>
<td>alpha</td>
<td>Cw</td>
<td>−0.07</td>
<td>−0.39</td>
<td>0.30</td>
<td>−0.41</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Lw</td>
<td>0.05</td>
<td>−0.26</td>
<td>0.40</td>
<td>0.33</td>
<td>0.74</td>
</tr>
</tbody>
</table>

bMCU post = posttest score in the basic Mechanics Conceptual Understanding test; Cw = mean clustering coefficient; Lw = characteristic path length; CI = confidence interval, est = estimate; est std = standardized estimate.
4.4 Discussion

We investigated the association between specific working memory functions and individual differences in conceptual learning in a sample of adolescents in the domain of physics. This approach widened the scope of previous research, which had a strong focus on mathematics learning in children (Peng et al., 2016; Raghubar et al., 2010). The predictive value of a theoretically driven selection of working memory tasks, specifically a color-word Stroop task, a visual arrays task, an updating task, and a secondary memory task, for conceptual learning in an educational context was examined. Additionally, we measured brain activity with EEG during working memory performance to examine whether neural measures uniquely inform the understanding of conceptual learning. The present study, to the best of our knowledge, is the first to investigate the association between these working memory functions and physics learning in a group of adolescents from a behavioral as well as a neural perspective.

With respect to the behavioral data, we ran a regression analysis and found an association with conceptual learning only for the secondary memory task. In this task, the participants remembered several locations within a matrix and had to recall them in serial order after solving an unrelated task (in this case, solving arithmetic equations). Theoretically, when attention is moved away from the list, e.g., due to the processing of the unrelated task, it is no longer actively maintained in primary memory and has to be retrieved from secondary memory (see Unsworth & Engle, 2007). Through the ability to narrowly constrain memory search with more specific retrieval cues, higher compared to lower WMC participants are assumed to perform better in both working memory and episodic memory tasks (Unsworth, 2009). The fact that especially the performance in this task, after controlling for the processes involved in the other tasks, is predictive of conceptual learning in an ecologically valid school setting is notable. Although the secondary memory task shares general working memory processes such as maintenance and rehearsal with other tasks, it is set apart by the explicit link with long-term memory, i.e., the requirement to move information from secondary memory back to primary memory. The importance of the retrieval of already available knowledge is in accordance with a major tenet of educational psychology stressing the role of prior knowledge for learning (e.g., Ackerman, 2007; Stern, 2015; Tricot & Sweller, 2013). Prior knowledge is particularly relevant for conceptual learning, which is highly
incremental (i.e., concepts are hierarchically organized) and reliant on radical knowl-
edge restructuring (Carey, 2000). For example, successful learning in the domain of
physics requires to substitute everyday explanations, such as assuming that a continuous
force must be applied to move an object, by scientific ones (Clement, 1982). This kind
of learning is sustainable if students are fully aware of the limitations of their former
knowledge. Therefore, everyday and scientific explanations have to be activated and
consciously compared in parallel before a scientifically appropriate knowledge structure
can be built. It is thus thoroughly plausible that the ability to recall only the relevant
information and dismiss irrelevant information might have helped in the present study
to build up more coherent conceptual knowledge throughout the instruction.

Although none of the other tasks had a unique effect on conceptual learning, it goes
without saying that other working memory processes play an important role. The ability
to stay on task and inhibit irrelevant information, indexed by the Stroop task, and the
maintenance and updating of information, indexed by the visual arrays and updating
tasks, are clearly also relevant. However, in our sample the variance in these tests did
not account for differences in conceptual learning. Future studies should investigate the
robustness and generalizability of the present findings with a larger and less selective
sample, ideally using multiple tasks per process to account for task-specific variance.
This would also allow for a better isolation of the specific cognitive functions relevant
for conceptual learning.

Another goal of the present study was to examine the added value of neuroimaging to
inform the association between working memory and conceptual learning. One typical
critique of neuroscientific studies is the fact that the experimental environment is highly
artificial, oftentimes relying on too simple and unrealistic material, which diminishes
the ecological validity and applicability of the findings to learning in classrooms (see
e.g. De Smedt et al., 2010). One way around this basic limitation has gained popularity
in recent years: The use of neural data in addition to behavioral data to predict future
performance or, with respect to an educational intervention, to predict which students
profit more or less from it. Most famously, Supekar et al. (2013) found that neural
measures were more sensitive than behavioral measures in predicting learning gains in
an eight-week math tutoring program for children (see also Dumontheil & Klingberg,
2012; Evans et al., 2015; Hoeft et al., 2011).

In this vein, we measured brain activity with EEG during working memory perfor-
4.4 Discussion

mance in the laboratory and related it to conceptual learning in classrooms. However, in contrast to earlier studies, we did not find any associations with either power or functional network characteristics after controlling for behavioral measures, such as working memory performance and reasoning ability. This suggests that, at least in the present study, brain activity during working memory performance did not provide added value for predicting conceptual learning. However, there are several alternative explanations for this: First, although the samples were of similar size in other studies (Dumontheil & Klingberg, 2012; Hoeft et al., 2011; Supekar et al., 2013), it cannot be denied that our sample was relatively small and that more participants might have been needed to reveal an actual effect. Second, it could be argued that we chose unsuited markers of neural activity. We believe, however, that the frequency bands we selected are the best researched and understood components of the human EEG (Antonenko et al., 2010; Klimesch et al., 2008). They were also highly sensitive to behavioral measures, in particular reasoning ability, in a related study with the same sample (Rütsche et al., 2017). Furthermore, the wPLI is a state-of-the-start connectivity index that addresses several critical problems of classical measures such as coherence and thus is generally more adequate to characterize functional brain networks (Vinck et al., 2011). Third, there was a delay between the present study and the end of classroom study, which could have obscured the association between brain measures and conceptual learning. However, the fact that we actually found an effect for behavioral measures suggests that this cannot be the sole reason.

Taken together, the present study had two mains goals: First, we wanted to investigate the relationship between specific working memory functions and ecologically valid conceptual learning in a sample of adolescents in the domain of physics. Second, we were interested whether neuroimaging increases our understanding of conceptual learning over and above behavioral measures. We found a unique relation between the secondary memory task, specifically requiring the retrieval from secondary memory or the activated portion of long-term memory, and conceptual learning. This ability might have helped to retrieve the relevant prior knowledge and, in turn, to build up a well-organized network of conceptual knowledge. Furthermore, in contrast to earlier studies, there was no added value of neural measures to predict educational outcomes in addition to behavioral performance. These findings highlight the relevance of a particular working memory function for conceptual learning in an educational context and suggest
that more work has to be done to flesh out opportunities and limits of neural measures to inform learning in an educational context.

4.5 Acknowledgments

We would like to thank Caroline Wölffe for her help during the EEG recording.

References


References


References


References


99
References


References


References


References


Unsworth, N. & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search...
References


5 Constructing and Refining Relational Categories: Unraveling the Interplay between Prior Knowledge, Reasoning Ability, Working Memory Capacity, and Strategy

Bruno Rütsche\textsuperscript{a}, Lennart Schalk\textsuperscript{a}

\textsuperscript{a}ETH Zurich, Research on Learning and Instruction

Abstract

A typical demand in modern education is incremental learning of relational categories and concepts: Learners have to construct a first understanding of a category and have to refine it across subsequent lessons. The multitude of processes and individual characteristics influencing learning outcomes in educational settings makes the identification of their unique contributions difficult. We unraveled the interplay between reasoning ability, working memory capacity, and strategic approaches (exemplar- versus rule-based) in a category learning paradigm. The paradigm allowed for the close supervision of relational knowledge construction and subsequent refinement by using specific learning and transfer tasks. While reasoning ability predicted knowledge construction, working memory capacity together with prior knowledge (i.e., the initially constructed knowledge) predicted knowledge refinement. Learners following rule-based strategies achieved higher learning gains in general. We additionally provide a series of explorative mediation analyses to further unravel the mechanisms underlying relational category learning. These finding illustrate the potential of incremental learning paradigms to serve as a bridge between basic and applied research on category and concept learning.
5.1 Introduction

*Keywords*: relational category learning; concepts; prior knowledge; reasoning ability; working memory capacity; strategy

*Author contributions*: BR and LS were both involved in the study design, acquisition and interpretation of data as well as in drafting and revising the manuscript. BR performed the statistical analysis.

### 5.1 Introduction

Learning relational categories and concepts is a central requirement of modern education (Resnick, 2010). Scientific concepts such as Newtonian laws in physics or evolution in biology are powerful categorization tools. They abstract from specific instances to provide generic descriptions capturing only relational characteristics (Goldwater & Schalk, 2016). For example, Newton’s third law denotes that a pair of forces acts in any interaction of two objects and evolutionary theory suggests that mutation and gene migration create variation in every species. This abstraction makes these concepts applicable to various phenomena which seem dissimilar at first glance (Chi & VanLehn, 2012). For example, a general understanding of the underlying physiological processes and principles of immune disorders helps medical doctors to correctly diagnose patients showing diverse patterns of symptoms. Typically, understanding of relational categories is gradually constructed over time requiring the continuous, incremental refinement of knowledge (Dixon & Kelley, 2007). These construction and refinement processes are challenging and learners differ in how they acquire relational category knowledge (Craig & Lewandowsky, 2012; McDaniel, Cahill, Robbins, & Wiener, 2014). We developed a category learning paradigm to unravel how interindividual characteristics influence the construction and refinement of relational categories.

#### 5.1.1 Category Learning and Conceptual Change

Category and concept learning are major research strands in basic cognitive and applied educational psychological research, but are not well integrated. Cognitive psychologists have developed sophisticated category learning paradigms to precisely monitor category acquisition and use. Generally, categories that are defined by a common set of (per-
5.1 Introduction

Conceptual features, so-called feature-based categories, are distinguished from categories characterized by a shared relational structure, so-called relational categories (Markman & Stilwell, 2001). Prior research has shown that feature-based categories differ from relational (or rule-based) categories both in terms of their representation and with regard to which cognitive processes are required or recruited during learning (Ashby & Maddox, 2005, 2011). While relational categories are of higher relevance for education, most research has focused on feature-based categories (Goldwater & Schalk, 2016). Moreover, the studies that investigated relational category learning typically scrutinize category construction but not further refinement and control the influence of prior knowledge by using artificial materials like arrangements of geometrical figures or alien species (Craig & Lewandowsky, 2012; Little & McDaniel, 2015a; McDaniel et al., 2014). However, educational psychologists who investigate conceptual change (Carey, 2000; Vosniadou, 2008), that is, the incremental learning of scientific concepts and categories, emphasize that the development and refinement of real-world relational concepts and categories strongly depends on prior knowledge (Ackerman, 2007; Schneider & Hardy, 2013; Stern, 2015; Tricot & Sweller, 2013). Thus, focusing on (feature-based) category construction and neglecting the influence of prior knowledge decreases the ecological validity of category learning research for education and impedes the integration of basic and applied psychological research.

At the same time, theoretical advances in conceptual change research have become increasingly difficult. The multitude of factors influencing learning in educational settings complicate pinpointing the specific contributions of cognitive mechanisms and processes. Goldwater and Schalk (2016) therefore suggested to bridge basic category learning and applied conceptual change research: The rigorous methodology of category learning research should be adapted to investigate the construction and refinement of relational categories, taking into account the influence of prior knowledge.

5.1.2 Interindividual Differences in Relational Category Learning

Recent research has shown that cognitive abilities like reasoning and working memory capacity (WMC) explain interindividual differences in relational category construction (Craig & Lewandowsky, 2012; Little & McDaniel, 2015a; McDaniel et al., 2014). Reasoning refers to the ability to solve abstract problems; it is typically measured by tasks
5.1 Introduction

requiring analogical and relational comparisons (e.g., Raven, Raven, & Cout, 2003). WMC refers to interindividual differences in the ability to access, hold, and manipulate information (e.g., letters, numbers, spatial positions); it is typically measured with a battery of computerized tasks (e.g., Lewandowsky, Oberauer, Yang, & Ecker, 2010). Although both constructs share common variance (about 50%), they are not isomorphic (Kane et al., 2004; Oberauer, Schulze, Wilhelm, & Süß, 2005). Cognitive abilities are generally positively correlated with success in category learning (Craig & Lewandowsky, 2012; Little & McDaniel, 2015a; McDaniel et al., 2014).

Individuals do also differ in the strategies they pursue in relational categorization tasks (Little & McDaniel, 2015a; McDaniel et al., 2014). McDaniel et al. (2014) showed that learners could be distinguished based on their transfer performance; that is, how they solved new, untrained cases in a function learning task. While some learners focused on abstracting the relational rule governing the category system during training and could transfer the rule to solve new cases; others focused on rote learning of specific exemplars during training (i.e., they memorized pairings of numerical input and output values) and consequently failed in solving new cases. Little and McDaniel (2015a) used ambiguous transfer items that made it possible to distinguish between learners relying on superficial or relational similarity in categorization. The ambiguous transfer items were perceptually similar to the training items (i.e., they overlapped in features), but their relational structure required classification in the opposing category. Consequently, categorizing based either on superficial or relational similarity resulted in different classifications. These classifications could by predicted by strategy self-reports (SSR) provided by the learners. Importantly, this difference in strategy adoption between individuals might also influence learning in educational contexts (Little & McDaniel, 2015b). In science, learners have to understand that problems are typically categorized based on their common relational structure rather than their superficial features (Chi & VanLehn, 2012). Taken together, cognitive abilities and strategies explain interindividual differences in relational category learning.

5.1.3 The Current Study

Recent research on knowledge restructuring provides a first step to bridge the gap between basic category learning and applied conceptual change research. Sewell and
5.1 Introduction

Lewandowsky (2011, 2012), for example, investigated how knowledge acquired within an experimental session could be restructured and applied for categorization (e.g., by coordinating partial rules). In contrast, the focus of the present study was on incremental learning of relational categories. To this aim, we developed a category learning paradigm that allowed for the monitoring of the construction and refinement of relational categories in a well-controlled environment.

The paradigm comprised two successive learning phases. Learners first constructed knowledge about two diseases that were defined by an XOR-relation between the values of two medical tests. Subsequently, they refined this knowledge (i.e., their prior knowledge) with respect to when treatment was required and when it was not required. This sequence mimics learning in educational settings, where a teacher introduces a new concept in a first lesson and refines it in a second lesson. After each learning phase, we assessed participants’ category knowledge with five distinct types of items: trained items as well as unambiguous, ambiguous, intra-range, and extra-range transfer items (for details see the Method section). Training items were identical to items that were presented during the learning phase. Unambiguous transfer items and ambiguous transfer item shared one test value with the training items, but either resulted in the same (unambiguous) or opposing (ambiguous) classification responses by learners following exemplar- and rule-based strategies. Intra-range and extra-range items shared no test value with trained items and were only reliably solvable with a rule-based strategy. To identify learners’ strategies, we used SSR as well as the pattern of performance on the transfer items. In a second experimental session, we collected reasoning ability and WMC as measures of cognitive ability. Bringing together the measures from both experimental sessions allowed us to unravel the effect of prior knowledge, cognitive abilities, and strategy and pinpoint their specific and shared effects in incremental category learning.

We expected that knowledge construction would be mainly predicted by reasoning ability (e.g. Little & McDaniel, 2015a; McDaniel et al., 2014). Typically, reasoning ability loses much of its impact as soon as the effect of prior knowledge is controlled (Ackerman, 2007; Lohman, 1999; Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2013; Staub & Stern, 2002; Stern, 2015; Tricot & Sweller, 2013; Weinert, Helmke, & Schneider, 1990). In accordance with the investment theory of intelligence (e.g., Kvist & Gustafsson, 2008; Schweizer & Koch, 2001), we assumed that this decrease might in-
5.2 Method

dicate that prior knowledge at least partly mediates the effect of reasoning ability on knowledge refinement. Hence, we expected that prior knowledge would be the most important predictor for knowledge refinement.

With regard to WMC, we hypothesized that it would be especially related to knowledge refinement. Knowledge refinement poses higher demands on WMC than knowledge construction because prior knowledge has to be retrieved, integrated with novel knowledge, and re-stored. In line with this assumption, empirical findings generally show that WMC still plays a role in learners with high levels of prior knowledge (Hambrick & Meinz, 2011).

Furthermore, based on previous research, we expected that reasoning ability and WMC would not be directly predictive of transfer performance or strategy (Craig & Lewandowsky, 2012). We also hypothesized that SSR would predict performance on items that can be solved reliably by rule-based strategies (i.e., ambiguous, intra-range, and extra-range items), but not on items that can be solved by either exemplar-based or rule-based strategies (i.e., unambiguous items; Little & McDaniel, 2015a). Moreover, we expected that following a rule-based strategy would be beneficial for learning, particularly in the performance-based measures of strategy due to their higher sensitivities (McDaniel et al., 2014).

5.2 Method

5.2.1 Participants

The sample size was determined by considering the effect sizes reported in recent research on individual differences in category learning (e.g., Craig & Lewandowsky, 2012; Little & McDaniel, 2015a). We tested 112 Swiss university students majoring in a variety of subjects in Zurich, Switzerland. Seven participants were excluded from statistical analysis: two due to technical problems during data collection; five did not attend the second session. The final sample consisted of 105 participants (63 females; $M_{age} = 22.47, SD_{age} = 2.23$). The institutional ethics review board approved the study; it was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. All participants gave written informed consent and received 80 CHF (about 80 USD) expense allowance.
5.2 Method

5.2.2 Procedure

The study consisted of two sessions with a duration of about 90 minutes each. In the first session, we administered the relational category learning paradigm and collected SSR. In the second session one week later, we assessed reasoning ability, WMC, and gathered information about the subjects (e.g., gender, age). Participants performed all tasks on computers in a group lab with an average group size of 26 participants per session.

5.2.3 Measures

Relational category learning paradigm

The paradigm consisted of a first learning phase in which participants constructed prior knowledge (construction phase), a first assessment in which we measured the acquired category knowledge (assessment 1), a second learning phase in which participants refined the knowledge (refinement phase), and a second assessment in which we again measured the acquired category knowledge (assessment 2). In the construction phase, we instructed participants to imagine traveling to a remote jungle region to replace a recently deceased medical doctor. In this region, the inhabitants suffered from one of two fictional diseases: Burlosis or Midosis (the names were borrowed from a classical category learning study by Medin, Altom, Edelson, & Freko, 1982). Participants learned to diagnose these diseases with the aid of old patient records left behind by their predecessor.

The diseases were defined by an XOR-relation between the two test results. For illustrative purposes, Figure 5.1 depicts this relation as a two-dimensional coordinate system: The values of Test B (y-axis) are plotted against those of Test A (x-axis). The origin of the system was arbitrarily set to Test A = 241 and Test B = 359; these values represented the cutoffs (i.e., category boundaries). According to the XOR-relation, patients suffered from Burlosis when either both test results were above or below the respective cutoffs (i.e., both values were either from quadrant I or III), while patients suffered from Midosis when one test result was above and the other below the respective cutoffs (i.e., both values were either from quadrant II or IV). Participants learned to diagnose the diseases across 10 blocks based on trial-by-trial feedback. In each block,
the same 16 training items were presented in random order. Participants were neither informed about the cutoffs nor were they shown an illustration of the category space as illustrated in Figure 5.1.

For each item, participants were shown the two test results and had to decide which disease was indicated by these results. Participants responded by clicking the respective button with the mouse (see Figure 5.2). Feedback about their decision was provided immediately afterwards: either “correct” or “wrong” was presented for 1000 ms. In the intertrial interval, a fixation point was shown for 1000 ms. At the end of each block, we included a self-timed pause. This pause was followed by an information screen showing the current and total number of training blocks for 5000 ms. We randomly assigned two sets of training items to the participants to control that effects were not caused by specific combinations of items (for details on the item generation see the Appendix).

After the construction phase, participants proceeded with assessment 1. They were told that new patients would come to their doctor’s office and that they would conduct medical Tests A and B to diagnose each patient. Assessment 1 included training items, which were identical to the items studied in the construction phase as a measure of learning, and various kinds of transfer items.

The transfer items were comprised of unambiguous, ambiguous, and intra-range items. Both unambiguous and ambiguous transfer items shared values with the training items. Specifically, for each training item, there was one unambiguous and one ambiguous item that had either an identical value in Test A or Test B. The dashed line in Figure 5.1 illustrates a case in which a training item, an unambiguous, and an ambiguous item share a Test B value of 375. For the unambiguous items, the other test value was selected from the same quadrant as the value of the corresponding training item (e.g., training: Test A = 279; unambiguous: Test A = 261). Thus, the superficial similarity (i.e., the identical value) of these transfer items to the training items correctly indicated category membership, resulting in the same response when classification was either based on an exemplar-based or a rule-based strategy. For the ambiguous items, the other test value was selected from the opposing quadrant (e.g., training: Test A = 279; ambiguous: Test A = 227), rendering them members of the other category. Accordingly, the responses to these items were expected to diverge for different types of strategies: Learners following an exemplar-based strategy should incorrectly classify the ambiguous items due to their misleading superficial similarity, while learners following a rule-based strategy
5.2 Method

Figure 5.1: Illustration of the relational category system. Each item consisted of a combination of a Test A (x-axis) and a Test B value (y-axis). These values were either above or below test-specific cutoff values (represented by the black lines). Training items were presented in the learning phases; transfer items (unambiguous, ambiguous, intra-range, extra-range) items were presented in assessment 1 and 2. The dashed line highlights an exemplary case of a trained item, its corresponding unambiguous transfer item (which shares one value with the trained item that was indicative for category membership), and its corresponding ambiguous transfer item (which shares one value with the trained item that was not indicative for category membership). For further details on the items see the Method section and the Appendix.
5.2 Method

Figure 5.2: Sample items for the construction phase and assessment 1 (left panel) as well as the refinement phase and assessment 2 (right panel). Participants responded by clicking on one of the dark gray buttons.

should be able to ignore the superficial similarity and correctly classify them. Intra-range items did not share any value with the training items; instead, they consisted of new test values from within the range of values studied in the learning phase (for details see the Appendix). These items could thus only reliably be solved correctly if learners abstracted the relational XOR-rule governing the category system.

In total, participants solved 64 items in assessment 1: 16 training items, 16 ambiguous, 16 unambiguous, and 16 intra-range transfer items (for details see the Appendix). All participants solved the same set of items, which were presented in 4 blocks of 16 items each. Each block contained one item per item type and quadrant. The task structure was identical to the construction phase, but no performance feedback was provided (see Figure 5.2). We refer to the training items in assessment 1 as Trained_1 and to the transfer items as Transfer_1.

After a short pause, participants proceeded with the refinement phase in which they had to revise their knowledge constructed in the first learning phase. The task instructions mentioned that additional records from their predecessor were found in an archive. These records indicated that even though all patients had either one of the two diseases, treatment was not always necessary. Thus, participants had to learn in which cases treatment was required: Patients with test results from quadrants I, II, or III needed treatment, while patients from quadrant IV (i.e., Midosis items with high Test A and
low Test B values) needed no treatment (see Figure 5.1). The task structure in the refinement phase was identical to the construction phase only that the participants now responded by clicking on one of four buttons to indicate the disease as well as the necessity of a treatment (see Figure 5.2).

Assessment 2 closely resembled assessment 1. The same cover story was used and the same 64 items were presented, but in a different order. Using the same items was possible because we provided no feedback in assessment 1; moreover, using the same items increases the comparability of both assessment phases. After presenting these transfer items, we additionally presented 16 extra-range transfer items (4 per quadrant). The extra-range items did not share any value with the training items and required the acquired knowledge to be applied to test results lying outside of the range studied in the learning phases (i.e., Test A: > 203 or > 279, Test B: < 321, > 397; see Figure 5.1). Thus, extra-range items – like the intra-range items – could only be reliably solved correctly if participants abstracted the relational XOR-rule and understood when treatment was necessary. We omitted extra-range items in assessment 1 because the increased range could have prompted learners to switch to a rule-based strategy for the subsequent refinement phase. The task structure of assessment 2 was identical to the refinement phase, but without feedback. We refer to the training items in assessment 2 as Trained_2 and to the transfer items as Transfer_2. After assessment 2, participants received feedback about their average performance in each phase and were informed about the correct categorization rules.

**Strategy self-reports (SSR)**

After each assessment, we collected SSR with questions adapted from Little and McDaniel (2015a). That is, we asked participants whether they focused on memorizing the values of the medical tests and the corresponding diseases during learning or whether they tried to develop a rule. Similarly, we asked whether they diagnosed the new patients in the assessments based on similarity of values seen for patients in the learning phases or whether they tried to apply a rule. Participants answered all questions on a Likert scale ranging from 1 to 7, with higher values indicating an increased reliance on rules.
5.3 Results

Cognitive abilities

We measured two cognitive abilities: reasoning ability and WMC (for details see the Appendix). For reasoning ability, we used a computerized version of the RAPM (Raven et al., 2003) contained in the Wiener Test System (https://www.schuhfried.at/). For WMC, we used the sentence span, memory updating, and spatial short-term memory tasks from the psychometrically validated task battery by Lewandowsky et al. (2010). However, the latent factor representing WMC in the statistical analysis was based on only the memory updating and sentence span tasks (for details and explanations see the Appendix).

![Figure 5.3](image)

Figure 5.3: Solution rates in both learning phases. Participants improved significantly over training blocks in the construction (left panel) and the refinement phase (right panel). Error bars indicate 95% confidence intervals.

5.3 Results

5.3.1 Learning Phases

First, we checked whether participants improved in their categorization performance over time in the construction and refinement phases (see Figure 5.3). We conducted repeated measures ANOVAs with the learning blocks as within-subject factor and compared the last and first block with paired t-tests. Participants successfully learned in the
construction phase \( F(5.54, 576.42) = 44.81, p < 0.001, \eta^2_G = 0.30 \), with large gains from the first to the last block \( t(104) = 11.44, p < 0.001, d = 1.53 \). Mean performance exceeded guess rate (50%) from block 3 onward (95%-CI [50.7 55.5]). Participants also successfully learned in the refinement phase \( F(6.82, 709.25) = 44.64, p < 0.001, \eta^2_G = 0.30 \), again with large gains from the first to the last block \( t(104) = 13.77, p < 0.001, d = 1.24 \). Performance at block 1 (95%-CI [39.2 44.8]) was between 25% (chance level without any prior knowledge about the disease classification) and 50% (chance level with perfect prior knowledge about the disease classification), suggesting that knowledge from the construction phase was successfully applied in the refinement phase. Thus, participants were able to construct and refine their knowledge about the category system.

5.3.2 Assessments 1 and 2

To test our hypotheses regarding the transfer performance, we ran separate structural equation models for unambiguous, ambiguous, intra-range, and extra-range transfer items. Table 5.1 compiles descriptive statistics and correlations between all observed measures. Figure 5.4 depicts the overall structure of the models and the results of the path analyses. Table 5.2 lists the direct effects and Table 5.3 lists the indirect effects for all models (for further details on the statistical analyses see the Appendix). We present the results in the following order. First, we provide measures of fit for the models. Second, we compare the effects of each predictor (prior knowledge as acquired in the construction phase, reasoning ability, WMC, and SSR) across models to integrate the findings. We provide standardized regression weights and highlight statistically significant effects \( p < 0.05 \) with an asterisk.

Model fits

We assessed the fit of the structural equation models via \( \chi^2 \), RMSEA, CFI, and SRMR, using non-significant \( \chi^2 \), RMSEAs smaller than 0.05 with a narrow 90% confidence interval including 0, CFI values larger than 0.97, and SRMR values smaller than 0.05 as indicators of good model fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003). The models for unambiguous items \( \chi^2(14) = 9.50, p = 0.798; \ RMSEA = 0, 90\%-CI [0, 0.06]; \ CFI = 1; \ SRMR = 0.03 \), ambiguous items \( \chi^2(14) = 13.25, p = 0.507; \ RMSEA
5.3 Results

Figure 5.4: Results of the structural equation models estimated for each type of transfer item. For each regression path, standardized coefficients as well as their significance (*) are given for each type of transfer item (format: $[\beta_{\text{unambiguous}}, \beta_{\text{ambiguous}}, \beta_{\text{intra-range}}, \beta_{\text{extra-range}}]$). MU = Memory Updating; SS = Sentence Span; WMC = Working Memory Capacity; SSR = Strategy Self-Report; Trained_1 and Transfer_1 = performance on training and transfer items in assessment 1; Trained_2 and Transfer_2 = performance on training and transfer items in assessment 2. Black arrows indicate significant regression paths (in at least one model) and significant covariances. Gray arrows indicate non-significant regression paths, non-significant covariances as well as factor loadings for WMC and SSR (for further details see the Results section and the Appendix).
Table 5.1: Descriptive statistics of all observed measures and their correlations.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>α</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reasoning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 RAPM</td>
<td>105</td>
<td>26.90</td>
<td>4.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WMC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 MU</td>
<td>100</td>
<td>61.62</td>
<td>17.00</td>
<td>0.91</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 SS</td>
<td>105</td>
<td>73.10</td>
<td>14.00</td>
<td>0.82</td>
<td>0.23</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 SSTM</td>
<td>105</td>
<td>86.41</td>
<td>5.24</td>
<td>0.82</td>
<td>0.36</td>
<td>0.23</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SSR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Learning</td>
<td>105</td>
<td>4.89</td>
<td>1.68</td>
<td>0.89</td>
<td>0.04</td>
<td>0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Test</td>
<td>105</td>
<td>4.52</td>
<td>1.92</td>
<td>0.87</td>
<td>0.14</td>
<td>0.08</td>
<td>0.05</td>
<td>0.10</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Assessment 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Trained</td>
<td>105</td>
<td>70.39</td>
<td>16.74</td>
<td>0.38</td>
<td>0.18</td>
<td>0.15</td>
<td>0.05</td>
<td>0.02</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Unambiguous</td>
<td>105</td>
<td>62.00</td>
<td>15.23</td>
<td>0.10</td>
<td>-0.08</td>
<td>0.07</td>
<td>-0.06</td>
<td>0.08</td>
<td>0.03</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Ambiguous</td>
<td>105</td>
<td>58.44</td>
<td>17.26</td>
<td>0.08</td>
<td>0.00</td>
<td>0.10</td>
<td>-0.01</td>
<td>0.27</td>
<td>0.29</td>
<td>0.16</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Intra-range</td>
<td>105</td>
<td>60.70</td>
<td>17.13</td>
<td>0.04</td>
<td>-0.01</td>
<td>0.17</td>
<td>-0.13</td>
<td>0.32</td>
<td>0.29</td>
<td>0.36</td>
<td>0.65</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Assessment 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Trained</td>
<td>105</td>
<td>66.67</td>
<td>20.78</td>
<td>0.26</td>
<td>0.33</td>
<td>0.21</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.67</td>
<td>0.30</td>
<td>0.19</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Unambiguous</td>
<td>105</td>
<td>50.87</td>
<td>21.18</td>
<td>0.06</td>
<td>-0.01</td>
<td>0.09</td>
<td>-0.08</td>
<td>0.12</td>
<td>0.13</td>
<td>0.38</td>
<td>0.59</td>
<td>0.43</td>
<td>0.52</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Ambiguous</td>
<td>105</td>
<td>46.44</td>
<td>22.55</td>
<td>0.09</td>
<td>-0.08</td>
<td>0.12</td>
<td>-0.02</td>
<td>0.29</td>
<td>0.28</td>
<td>0.38</td>
<td>0.52</td>
<td>0.72</td>
<td>0.66</td>
<td>0.37</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Intra-range</td>
<td>105</td>
<td>50.98</td>
<td>22.74</td>
<td>0.04</td>
<td>0.00</td>
<td>0.13</td>
<td>-0.12</td>
<td>0.33</td>
<td>0.29</td>
<td>0.37</td>
<td>0.50</td>
<td>0.52</td>
<td>0.62</td>
<td>0.49</td>
<td>0.79</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Extra-range</td>
<td>105</td>
<td>40.50</td>
<td>21.30</td>
<td>0.15</td>
<td>0.07</td>
<td>0.16</td>
<td>0.03</td>
<td>0.30</td>
<td>0.34</td>
<td>0.46</td>
<td>0.48</td>
<td>0.51</td>
<td>0.53</td>
<td>0.42</td>
<td>0.71</td>
<td>0.70</td>
<td>0.72</td>
<td></td>
</tr>
</tbody>
</table>

RAPM = Raven’s Advanced Progressive Matrices; WMC = Working Memory Capacity; MU = Memory Updating; SS = Sentence Span; SSTM = Spatial Short-Term Memory; SSR = Strategy Self-Reports. α denotes Cronbach’s Alpha.
5.3 Results

= 0, 90%-CI [0, 0.09]; CFI = 1; SRMR = 0.04), intra-range items ($\chi^2(14) = 9.88, p = 0.771; \text{RMSEA} = 0, 90\%-\text{CI} [0, 0.07]; \text{CFI} = 1; \text{SRMR} = 0.04), and extra-range items ($\chi^2(14) = 15.50, p = 0.345; \text{RMSEA} = 0.03, 90\%-\text{CI} [0, 0.1]; \text{CFI} = 1; \text{SRMR} = 0.04) all provided a good fit to the data.

Prior Knowledge and Transfer Effects

The relations between participants’ performance on training and transfer items in assessment 1 and 2 (Trained_1, Transfer_1, Trained_2, and Transfer_2, respectively) can be summarized as follows. With the exception for the model for the ambiguous transfer items (see below), all paths were positive and significant.

First, successful learning in the construction phase (Trained_1) was related to successful transfer performance (Transfer_1; $\beta_{\text{unambiguous}} = 0.32^*, \beta_{\text{ambiguous}} = 0.15, \beta_{\text{intra-range}} = 0.40^*, \beta_{\text{extra-range}} = 0.40^*$) and successful knowledge refinement (Trained_2; $\beta_{\text{unambiguous}} = 0.62^*, \beta_{\text{ambiguous}} = 0.65^*, \beta_{\text{intra-range}} = 0.59^*, \beta_{\text{extra-range}} = 0.59^*$). Second, a rule-based compared to an exemplar-based strategy (Transfer_1) facilitated knowledge refinement (Trained_2; $\beta_{\text{unambiguous}} = 0.15^*, \beta_{\text{ambiguous}} = 0.12^*, \beta_{\text{intra-range}} = 0.21^*, \beta_{\text{extra-range}} = 0.21^*$). This effect was particularly evident in the models for the intra-range and extra-range items, since these shared no superficial similarity with the training items and were thus only solvable with a rule-based strategy. Transfer_1 also predicted Transfer_2 ($\beta_{\text{unambiguous}} = 0.44^*, \beta_{\text{ambiguous}} = 0.62^*, \beta_{\text{intra-range}} = 0.40^*, \beta_{\text{extra-range}} = 0.34^*$). Thus, the adopted strategies generally persisted across the knowledge construction and refinement phases. Third, successful learning in the refinement phase (Trained_2) was related to successful transfer performance in assessment 2 (Transfer_2; $\beta_{\text{unambiguous}} = 0.46^*, \beta_{\text{ambiguous}} = 0.31^*, \beta_{\text{intra-range}} = 0.39^*, \beta_{\text{extra-range}} = 0.31^*$).

As noted above, these relations were different in the model for ambiguous items. We anticipated this difference because it reflects the specific properties of the ambiguous items. First, the effect of Trained_1 $\rightarrow$ Transfer_1 was weaker ($\beta_{\text{ambiguous}} = 0.15$) than in the models for the other transfer items. We expected that the superficial similarity of the ambiguous transfer items to the training items (i.e., the shared value on one of the medical tests) should reduce performance for learners adopting an exemplar-based strategy, but should not affect learners adopting a rule-based strategy. To check this expectation, we computed correlations between Trained_1 and Transfer_1 separately.
5.3 Results

for learners primarily relying on an exemplar-based or a rule-based strategy. We defined these groups by mean SSR values (on a 7-point Likert scale with higher values indicating an increased reliance on rules): Learners with a mean value smaller than 3 were classified as exemplar-based learners and learners with a mean value larger than 5 as rule-based learners. Indeed, we found no significant correlation for exemplar-based learners ($r(20) = -.07$, 95%-CI [-.50 .39]), but a positive correlation for rule-based learners ($r(40) = .31^*$, 95%-CI [0.02 0.55]).

Second, in contrast to the other models, the Transfer_1 $\rightarrow$ Trained_2 path was non-significant in the model for ambiguous items ($\beta_{\text{ambiguous}} = 0.12$). This finding again follows from the specific properties of the ambiguous items. That is, we would not expect a correlation between Transfer_1 and Trained_2 within the exemplar-based learners, since this strategy is detrimental for Transfer_1 (as seen before) but a viable strategy for Trained_2. In contrast, a rule-based strategy should be equally effective for both Transfer_1 and Trained_2. Indeed, we found no correlation between these measures within exemplar-based learners ($r(20) = -.03$, 95%-CI [-.47 .42]), but a positive correlation within rule-based learners ($r(40) = .33^*$, 95%-CI [.05 .57]). These diverging relations in the model for ambiguous items provide additional support for the notion that a rule-based compared to an exemplar-based strategy is beneficial for learning and transfer.

Reasoning Ability (RAPM)

We found a positive direct effect of reasoning ability on Trained_1 ($\beta_{\text{unambiguous}} = 0.38^*$, $\beta_{\text{ambiguous}} = 0.38^*$, $\beta_{\text{intra-range}} = 0.38^*$, $\beta_{\text{extra-range}} = 0.38^*$) but no direct effect on Transfer_1 ($\beta_{\text{unambiguous}} = 0.05$, $\beta_{\text{ambiguous}} = 0.02$, $\beta_{\text{intra-range}} = -0.12$, $\beta_{\text{extra-range}} = -0.12$), Trained_2 ($\beta_{\text{unambiguous}} = -0.12$, -0.11, $\beta_{\text{intra-range}} = -0.08$, $\beta_{\text{extra-range}} = -0.09$), and Transfer_2 ($\beta_{\text{unambiguous}} = -0.07$, $\beta_{\text{ambiguous}} = 0.04$, $\beta_{\text{intra-range}} = -0.05$, $\beta_{\text{extra-range}} = 0.06$). Reasoning ability was related to better learning in the construction phase, while it did not directly influence knowledge refinement or transfer performance.

However, investigating the indirect effects, reveals a more complex picture (see Table 5.3). All indirect effects beginning with the Reasoning $\rightarrow$ Trained_1 path were significant in the models for unambiguous, intra-range, and extra-range items. In the model for ambiguous items, however, only the paths that did not include the (non-significant)
Table 5.2: Direct effects in the structural equation models for each type of transfer item.

<table>
<thead>
<tr>
<th>Regression coefficients</th>
<th>Unambiguous</th>
<th>Ambiguous</th>
<th>Intra-range</th>
<th>Extra-range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est</td>
<td>est std</td>
<td>95% CI</td>
<td>est</td>
</tr>
<tr>
<td>Trained_1 → Transfer_1</td>
<td>0.29</td>
<td>0.32</td>
<td>[0.12, 0.46]</td>
<td>0.16</td>
</tr>
<tr>
<td>Trained_1 → Transfer_2</td>
<td>0.77</td>
<td>0.62</td>
<td>[0.57, 1.00]</td>
<td>0.81</td>
</tr>
<tr>
<td>Transfer_1 → Transfer_2</td>
<td>0.20</td>
<td>0.15</td>
<td>[0.01, 0.41]</td>
<td>0.14</td>
</tr>
<tr>
<td>Transfer_1 → Transfer_2</td>
<td>0.61</td>
<td>0.44</td>
<td>[0.40, 0.82]</td>
<td>0.81</td>
</tr>
<tr>
<td>Trained_2 → Transfer_2</td>
<td>0.46</td>
<td>0.46</td>
<td>[0.32, 0.61]</td>
<td>0.33</td>
</tr>
<tr>
<td>Reasoning → Trained_1</td>
<td>1.52</td>
<td>0.38</td>
<td>[0.58, 2.25]</td>
<td>1.53</td>
</tr>
<tr>
<td>Reasoning → Transfer_1</td>
<td>0.17</td>
<td>0.05</td>
<td>[0.07, 0.94]</td>
<td>0.07</td>
</tr>
<tr>
<td>Reasoning → Transfer_2</td>
<td>-0.58</td>
<td>-0.12</td>
<td>[-1.71, 0.33]</td>
<td>-0.57</td>
</tr>
<tr>
<td>Reasoning → Transfer_2</td>
<td>-0.37</td>
<td>-0.07</td>
<td>[-1.12, 0.31]</td>
<td>0.22</td>
</tr>
<tr>
<td>WMC → Trained_1</td>
<td>0.62</td>
<td>0.10</td>
<td>[0.49, 0.96]</td>
<td>0.01</td>
</tr>
<tr>
<td>WMC → Transfer_1</td>
<td>-0.14</td>
<td>-0.15</td>
<td>[0.38, 0.08]</td>
<td>-0.04</td>
</tr>
<tr>
<td>WMC → Transfer_1</td>
<td>0.34</td>
<td>0.28</td>
<td>[0.14, 0.76]</td>
<td>0.33</td>
</tr>
<tr>
<td>WMC → Transfer_2</td>
<td>-0.12</td>
<td>-0.09</td>
<td>[0.36, 0.06]</td>
<td>-0.25</td>
</tr>
<tr>
<td>SSR → Trained_1</td>
<td>0.15</td>
<td>0.01</td>
<td>[-1.77, 2.04]</td>
<td>0.18</td>
</tr>
<tr>
<td>SSR → Transfer_1</td>
<td>0.66</td>
<td>0.07</td>
<td>[-1.34, 2.56]</td>
<td>3.06</td>
</tr>
<tr>
<td>SSR → Transfer_2</td>
<td>-0.97</td>
<td>-0.08</td>
<td>[-2.91, 0.71]</td>
<td>-1.29</td>
</tr>
<tr>
<td>SSR → Transfer_2</td>
<td>1.53</td>
<td>0.12</td>
<td>[0.43, 3.62]</td>
<td>2.03</td>
</tr>
</tbody>
</table>

Covariances

| Reasoning ↔ WMC | 30.41 | 0.44 | [18.49, 44.55] | 30.19 | 0.44 | [18.12, 44.25] | 30.26 | 0.44 | [18.16, 44.33] | 30.34 | 0.44 | [18.30, 44.65] |
| Reasoning ↔ SSR | 0.41 | 0.06 | [0.39, 0.76] | 0.42 | 0.06 | [0.90, 1.78] | 0.40 | 0.06 | [0.94, 0.16] | 0.44 | 0.07 | [0.89, 1.18] |
| WMC ↔ SSR      | 0.93 | 0.04 | [0.13, 1.76] | 0.88 | 0.03 | [0.43, 1.32] | 0.84 | 0.03 | [0.43, 1.32] | 0.91 | 0.03 | [0.43, 1.32] |

$R^2$

| Trained_1 | 0.15 | 0.15 | 0.15 |
| Transfer_1 | 0.12 | 0.11 | 0.25 |
| Trained_2 | 0.53 | 0.52 | 0.54 |
| Transfer_2 | 0.51 | 0.62 | 0.51 |

Asterisks (*) indicate significant effects. RAPM = Raven’s Advanced Progressive Matrices; WMC = Working Memory Capacity; SSR = Strategy Self-Reports; Trained_1 and Transfer_1 = performance on trained and transfer items in assessment 1; Trained_2 and Transfer_2 = performance on trained and transfer items in assessment 2; est = estimate; est std = standardized estimate; CI = confidence interval.
Table 5.3: Indirect effects in the structural equation models conducted for each type of transfer item.

<table>
<thead>
<tr>
<th>Path</th>
<th>Unambiguous</th>
<th></th>
<th>Ambiguous</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>est</td>
<td>est std</td>
<td>95% CI</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Transfer_1</td>
<td>0.44</td>
<td>0.12</td>
<td>[0.15,0.87]</td>
<td>0.24</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Trained_2</td>
<td>1.17</td>
<td>0.23</td>
<td>[0.50,1.98]</td>
<td>1.24</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Transfer_1 → Trained_2</td>
<td>0.09</td>
<td>0.02</td>
<td>[0.01,0.30]</td>
<td>0.03</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Trained_2 → Transfer_2</td>
<td>0.54</td>
<td>0.11</td>
<td>[0.23,1.02]</td>
<td>0.41</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Transfer_1 → Transfer_2</td>
<td>0.27</td>
<td>0.05</td>
<td>[0.09,0.62]</td>
<td>0.19</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Transfer_1 → Trained_2 → Transfer_2</td>
<td>0.04</td>
<td>0.01</td>
<td>[0.01,0.14]</td>
<td>0.01</td>
</tr>
<tr>
<td>SSR → Transfer_1 → Trained_2</td>
<td>0.13</td>
<td>0.01</td>
<td>[0.21,0.84]</td>
<td>0.44</td>
</tr>
<tr>
<td>SSR → Transfer_1 → Transfer_2</td>
<td>0.40</td>
<td>0.03</td>
<td>[0.79,1.70]</td>
<td>2.49</td>
</tr>
<tr>
<td>SSR → Transfer_1 → Trained_2 → Transfer_2</td>
<td>0.06</td>
<td>0.01</td>
<td>[0.09,0.38]</td>
<td>0.14</td>
</tr>
<tr>
<td>WMC → Trained_2 → Transfer_2</td>
<td>0.16</td>
<td>0.13</td>
<td>[0.06,0.41]</td>
<td>0.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Path</th>
<th>Intra-range</th>
<th></th>
<th>Extra-range</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>est</td>
<td>est std</td>
<td>95% CI</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Transfer_1</td>
<td>0.62</td>
<td>0.15</td>
<td>[0.24,1.17]</td>
<td>0.62</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Trained_2</td>
<td>1.11</td>
<td>0.22</td>
<td>[0.47,1.88]</td>
<td>1.11</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Transfer_1 → Trained_2</td>
<td>0.16</td>
<td>0.03</td>
<td>[0.03,0.42]</td>
<td>0.16</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Trained_2 → Transfer_2</td>
<td>0.47</td>
<td>0.09</td>
<td>[0.20,0.90]</td>
<td>0.35</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Trained_2 → Transfer_2</td>
<td>0.34</td>
<td>0.06</td>
<td>[0.12,0.73]</td>
<td>0.26</td>
</tr>
<tr>
<td>Reasoning → Trained_1 → Transfer_1 → Trained_2 → Transfer_2</td>
<td>0.07</td>
<td>0.01</td>
<td>[0.02,0.21]</td>
<td>0.05</td>
</tr>
<tr>
<td>SSR → Transfer_1 → Trained_2</td>
<td>0.88</td>
<td>0.07</td>
<td>[0.23,2.05]</td>
<td>0.89</td>
</tr>
<tr>
<td>SSR → Transfer_1 → Transfer_2</td>
<td>1.87</td>
<td>0.13</td>
<td>[0.80,3.46]</td>
<td>1.49</td>
</tr>
<tr>
<td>SSR → Transfer_1 → Trained_2 → Transfer_2</td>
<td>0.38</td>
<td>0.03</td>
<td>[0.10,0.98]</td>
<td>0.28</td>
</tr>
<tr>
<td>WMC → Trained_2 → Transfer_2</td>
<td>0.14</td>
<td>0.10</td>
<td>[0.05,0.35]</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Asterisks (*) indicate significant effects. RAPM = Raven’s Advances Progress Matrices; WMC = Working Memory Capacity; SSR = Strategy Self-Reports; Trained_1 and Transfer_1 = performance on trained and transfer items in assessment 1; Trained_2 and Transfer_2 = performance on trained and transfer items in assessment 2; est = estimate; est std = standardized estimate; CI = confidence interval.
5.3 Results

Trained_1 → Transfer_1 path were significant; again reflecting the specific structure of the ambiguous items as explained before. Thus, the influence of reasoning ability on Transfer_1, Trained_2, and Transfer_2 was mediated by its initial effect on Trained_1. Notably, not only knowledge refinement, but also transfer performance was indirectly affected by reasoning ability.

**Working Memory Capacity (WMC)**

In all models, WMC was not related to Trained_1 ($\beta_{\text{unambiguous}} = 0.02$, $\beta_{\text{ambiguous}} = 0.01$, $\beta_{\text{intra-range}} = 0.01$, $\beta_{\text{extra-range}} = 0.01$), while it had an effect on Trained_2 ($\beta_{\text{unambiguous}} = 0.28^*$, $\beta_{\text{ambiguous}} = 0.26^*$, $\beta_{\text{intra-range}} = 0.26^*$, $\beta_{\text{extra-range}} = 0.26^*$). Thus, while reasoning ability was associated with better knowledge construction, WMC improved knowledge refinement. Furthermore, WMC was neither related to Transfer_1 ($\beta_{\text{unambiguous}} = -0.15$, $\beta_{\text{ambiguous}} = -0.04$, $\beta_{\text{intra-range}} = -0.02$, $\beta_{\text{extra-range}} = -0.02$) nor to Transfer_2 ($\beta_{\text{unambiguous}} = -0.09$, $\beta_{\text{ambiguous}} = -0.19$, $\beta_{\text{intra-range}} = -0.10$, $\beta_{\text{extra-range}} = -0.06$), with the exception of a negative association between WMC and Transfer_2 in the model for ambiguous items (i.e., $\beta_{\text{ambiguous}} = -0.19^*$). However, due to the small size and unexpectedness of this effect, we refrain from any interpretation. Furthermore, the indirect effect of WMC → Trained_2 → Transfer_2 ($\beta_{\text{unambiguous}} = 0.13^*$, $\beta_{\text{ambiguous}} = 0.08^*$, $\beta_{\text{intra-range}} = 0.10^*$, $\beta_{\text{extra-range}} = 0.08^*$) was significant in all models. Thus, WMC influenced transfer performance (Transfer_2) indirectly via its positive effect on knowledge refinement (Trained_2).

**Strategy Self-Reports (SSR)**

SSR was neither associated with Trained_1 ($\beta_{\text{unambiguous}} = 0.01$, $\beta_{\text{ambiguous}} = 0.02$, $\beta_{\text{intra-range}} = 0.01$, $\beta_{\text{extra-range}} = 0.02$) nor with Trained_2 ($\beta_{\text{unambiguous}} = -0.08$, $\beta_{\text{ambiguous}} = -0.10$, $\beta_{\text{intra-range}} = -0.13$, $\beta_{\text{extra-range}} = -0.14$). This indicates that self-reports of exemplar- and rule-based strategies were not directly related to learning. However, as expected, both strategies differently affected transfer performance. Specifically, SSR significantly predicted Transfer_1 in all but the model for unambiguous items ($\beta_{\text{unambiguous}} = 0.07$, $\beta_{\text{ambiguous}} = 0.29^*$, $\beta_{\text{intra-range}} = 0.33^*$, $\beta_{\text{extra-range}} = 0.33^*$) and predicted Transfer_2 in the models for intra-range and extra-range items ($\beta_{\text{unambiguous}} = 0.12$, $\beta_{\text{ambiguous}} = 0.15$, $\beta_{\text{intra-range}} = 0.22^*$, $\beta_{\text{extra-range}} = 0.22^*$). These effects confirm our hypothesis that a rule-compared to an exemplar-based strategy was more beneficial for solving ambiguous,
5.4 Discussion

The acquisition and incremental refinement of relational knowledge lies at the core of today’s education. In realistic educational settings (e.g., in classrooms), the complex interplay of various factors and mechanisms on learning is hard to unravel. We developed a category learning paradigm that provided a first approximation for studying incremental learning across two subsequent lessons: Participants constructed prior knowledge in a first learning phase and refined it in a second learning phase. We additionally assessed reasoning ability, WMC, and strategy (exemplar- vs. rule-based strategies). This allowed us to unravel the interplay of these interindividual differences in relational category construction and refinement.

The impact of reasoning ability and WMC on learning depended on the learning phase. Reasoning ability directly predicted initial knowledge construction, but had only an indirect effect on subsequent knowledge refinement. This finding supports the investment theory of intelligence (e.g., Schweizer & Koch, 2001) and empirically illustrates the decreasing impact of cognitive abilities on predicting learning when prior knowledge is accounted for (Ackerman, 2007; Murayama et al., 2013; Staub & Stern, 2002; Stern, 2015; Tricot & Sweller, 2013; Weinert et al., 1990). Furthermore, as expected, WMC

intra-range, and extra-range items, while both strategies are equally effective for solving unambiguous items.

However, the lack of a significant effect of SSR on ambiguous items in Transfer_2 is not in line with this reasoning ($\beta_{\text{ambiguous}} = 0.15$). We assume that this deviation emerged because not enough residual variance was left after accounting for the significant indirect effect of SSR $\rightarrow$ Transfer_1 $\rightarrow$ Transfer_2 ($\beta_{\text{unambiguous}} = 0.03$, $\beta_{\text{ambiguous}} = 0.18^*$, $\beta_{\text{intra-range}} = 0.13^*$, $\beta_{\text{extra-range}} = 0.11^*$). The remaining indirect effects including the SSR $\rightarrow$ Transfer_1 path follow logically from the direct effects described before: For the model for unambiguous items, no effects were present due to the non-significant SSR $\rightarrow$ Transfer_1 path; for the model for ambiguous items, no effects involving the non-significant Transfer_1 $\rightarrow$ Trained_2 path were present; for the models for intra-range and extra-range items, all effects were present. Thus, although SSR does not directly affect knowledge construction, it indirectly influenced knowledge refinement via its effect on Transfer_1.
5.4 Discussion

directly predicted knowledge refinement, but not knowledge construction. In line with findings from Hambrick and Meinz (2011), this shows that WMC matters even when learners already have high domain-specific knowledge, arguably due to the increased cognitive demands to retrieve, integrate and re-store knowledge during knowledge refinement.

Consistent with findings from educational studies, prior knowledge was the most important predictor for knowledge refinement, outweighing the effect of reasoning ability and WMC (e.g., Ackerman, 2007; Tricot & Sweller, 2013). Thus, knowledge refinement was impaired when no good understanding was acquired in the first learning phase. This reinstates a major claim of educational researchers that considering interindividual differences in prior knowledge during instruction is of paramount importance for successful learning.

Reinforcing prior research about the distinction between exemplar and rule-based strategies (Little & McDaniel, 2015a; McDaniel et al., 2014), participants’ strategy was distinctly related to transfer performance. We had two indicators to identify strategies: SSR and the performance on the various types of transfer items. SSR predicted performance in rule-favored transfer items (i.e., ambiguous, intra-range, and extra-range items), but did not predict performance in training and unambiguous transfer items for which both strategies are viable. If strategy was identified via performance on the transfer items, the rule-based strategy also predicted learning. Performance-based measures of strategy thus seem more sensitive than those based on self-reports. Overall, our findings highlight how exemplar- and rule-based strategies impact transfer performance. In modern education, learners have to acquire abstract concepts such as Newtonian force or evolution that can be flexibly applied, i.e. transferred, to a diversity of superficially different phenomena and situations (Chi & VanLehn, 2012). Our study together with several other recent studies indicate that learners differ in their learning strategy (Little & McDaniel, 2015a; McDaniel et al., 2014). Thus, prompting learners to focus on a rule-based strategy when learning relational categories and concepts could be a valuable instructional assistance. However, more research is needed to evaluate whether the distinction between exemplar- and rule-based strategies matters in realistic educational settings and whether these strategies are stable within a learner or depend on the context and the learning task (Little & McDaniel, 2015b).

Category construction and refinement depended on a complex interplay of factors.
5.5 Acknowledgments

Reasoning ability and WMC predicted knowledge construction and refinement and this in turn indirectly predicted transfer performance. We also found that the adoption of an exemplar- or a rule-based strategy (resulting in a more or less successful transfer performance) was independent of cognitive abilities (see also Craig & Lewandowsky, 2012; Little & McDaniel, 2015a). However, participants who reported a stronger tendency to rely on a rule-based strategy showed a better transfer performance after initial knowledge construction and this also indirectly supported their subsequent knowledge refinement. These effects, although explorative and generally rather small, demonstrate the importance to simultaneously consider various interindividual differences and several learning and transfer measures when studying incremental learning of categories and concepts.

In educational settings, learners gradually build up complex relational knowledge systems over time. Goldwater and Schalk (2016) suggested that research on relational category learning could bridge between more basic cognitive category learning research and more applied educational conceptual change research. The present study provided a step in this direction. The newly developed category learning paradigm allowed us to investigate the initial stages of this incremental learning process in a well-controlled environment. We have provided a fine-grained analysis of how reasoning ability, WMC, and strategy influence the construction of relational knowledge, and unraveled how these individual characteristics interplay with prior knowledge during knowledge refinement. The precise statistical modeling of the results offers a basis to further psychological theories about incremental learning.

5.5 Acknowledgments

We thank Elsbeth Stern and Micah Goldwater for comments on a draft of this manuscript. Furthermore, we want to thank Nicole Bruegger, Andrina Castione, Tamara Krummenacher, and Yvonne Oberholzer for their assistance in testing the learning paradigm.
Appendix

Relational Category Learning Paradigm: Item Creation

In the following, we provide a step-by-step description of how the items for the relational category learning paradigm were constructed (see Figure 5.1 for an illustration of the category system). The cutoffs for the Test A and Test B dimensions were arbitrarily set to 241 and 359, respectively. Test A had a minimum value of 203 and a maximum value of 279; Test B had a minimum value of 321 and a maximum value of 397. Generally, we first created the training items for a single quadrant (e.g., quadrant I) and imposed several restrictions during the process.

First, for this quadrant, we fixed one medical test result (Test A or Test B) per item to either the most extreme or the second most extreme value in a dimension to emphasize the boundaries of the category space. Furthermore, we ensured that there were no highly indicative items with extreme values in both dimensions. For example, item 1 was assigned the highest value in one dimension (Test A = 279), while item 2 was assigned the second-lowest value in the same dimension (Test A = 243; see Figure 5.1). Conversely, item 3 was assigned the lowest value in the other dimension (Test B = 350), while item 4 was assigned the second-highest value in that dimension (Test B = 396).

Second, to ensure that the items covered the complete category space, the free values were only allowed to take on only values that were spaced apart in steps of 7 starting from the second-lowest to the second-highest value in each dimension (i.e., Test A: 250, 257, 264, 271; Test B: 368, 375, 382, 389). The items were completed by randomly sampling from this pool of values.

Third, to prevent the application of a parity rule for classification, we also controlled that each dimension contained exactly two odd and two even numbers across all the items. Steps 1 to 3 were performed twice to create two sets of training items.

Fourth, after selecting the training items, we proceeded with creating the transfer items in quadrant I. The unambiguous items were constructed by copying the training items and randomly replacing the fixed values with free values from the same quadrant (i.e., the non-fixed values of the unambiguous items matched the values of the training items). For example, in item 1, the fixed value (Test A = 279) was exchanged with 261, while the non-fixed value (Test B = 375) was retained. The procedure was the same for the ambiguous items, with the exception that the new values were selected...
from the opposite quadrant. For example, the corresponding ambiguous item to item 1 was created by exchanging the fixed value with Test A = 227. Intra-range and extra-range items were both constructed by randomly selecting unused values from the same quadrant as the training items. The intra-range items, however, consisted of values sampled from within the training range, while the extra-range items consisted of values sampled from outside the training range (i.e., Test A > 279, Test B > 397).

Fifth, after having selected both training and transfer items for quadrant I, we built the whole category system by rotation or reflection and subsequent translation of these items. Specifically, quadrant II was created by mirroring quadrant I on the plane y = 378.5, quadrant III was created by a 90° clockwise rotation of quadrant I, and quadrant IV was constructed by a 90° counterclockwise rotation of quadrant II. This procedure guaranteed that the average difference of both test results across the items was the same in each quadrant. Sixth, we checked the complete set of items again to guarantee that no value was used twice, with the exception of the explicitly introduced duplicates in the training and unambiguous/ambiguous items. Finally, we added a restriction that prevented clustering and ensured that the items were well-distributed. The paradigm was programmed with PsychoPy v1.81.00 (Peirce, 2007, 2009).

Reasoning Ability: Raven’s Advanced Progressive Matrices (RAPM)

We used a computerized version of the RAPM (Raven et al., 2003) contained in the Wiener Test System (https://www.schuhfried.at/) to assess reasoning ability. We presented Set I (12 problems) without time constraint and Set II (36 problems) with a standard administration time of 40 min. Each problem in the RAPM consists of a 3×3 array of figures containing several elements such as lines and textures. One of the 9 figures in the array is missing and has to be filled with 1 of 8 response alternatives.

Working Memory Capacity (WMC)

To assess WMC, we used three working memory tasks (i.e., sentence span, memory updating, spatial short-term memory) from the psychometrically validated MATLAB task battery developed by Lewandowsky et al. (2010). The advantage of using a battery of heterogeneous tasks to assess WMC is that it reduces the contribution of task-specific variance, resulting in a more pure measurement of WMC. We presented the same ran-
domly generated sequence of stimuli and the same order of tasks to all participants. Unless otherwise stated, we used the default settings for the tasks.

In the sentence span task, the participants were alternatingly presented with sentences (e.g., “Each house has a roof”) and single letters. Their task was to judge the meaningfulness of the sentences (i.e., right arrow button: true, left arrow button: false) and to remember the letters. After processing 4 to 8 sentences and letters, the letters had to be recalled in the order of presentation. We replaced the original English sentences with similarly structured German sentences to adapt the task to our sample.

In the memory updating task, the participants saw numbers within 3 to 5 frames on the screen. These numbers were then updated several times based on simple arithmetic operations presented within the respective frame (e.g., +1). At the end of a trial, the final number contained in each frame had to be recalled.

In the spatial short-term memory task, the participants were required to remember 2 to 6 sequentially presented dots displayed within a 10×10 grid. They were instructed to focus on the overall pattern and not the absolute positions. At the end of each trial, the pattern had to be reproduced by clicking in the respective fields of the grid. Compared with the original version, we reduced the number of trials from 30 (6 trials per set size) to 20 (4 trials per set size) to save time.

**Statistical Analyses**

We first assessed whether participants improved in their categorization performance in both learning phases. Therefore, we computed the solution rates in each block and performed repeated measures ANOVAs, providing Greenhouse-Geisser corrected degrees of freedom and $p$-values if sphericity was violated.

To test our hypotheses regarding the transfer items, we ran separate structural equation models for unambiguous, ambiguous, intra-range and extra-range transfer items in Mplus (Version 7.11) using Full Information Maximum Likelihood (FIML) estimation (Muthén & Muthén, 2010). Structural equation modeling (SEM) has several advantages over conventional multiple regression analysis. SEM allows to go beyond observed variables and extract measurement-error-free latent variables, it allows to combine several regressions into a larger more coherent model, and it provides sophisticated indices to assess model fit as well as a high degree of flexibility in model specification. Figure 5.4
Appendix

depicts the estimated structural equation models: rectangles represent observed variables, circles represent latent variables, double-headed arrows represent covariances, and single-headed arrows represent factor loadings or regression weights. Table 5.1 shows descriptive statistics and the correlations between all observed measures.

The structure of the generic model can be summarized as follows (see Figure 5.4). First, the sequence of the variables Trained_1, Transfer_1, Trained_2, and Transfer_2 reflected the order of events in the relational category learning task (i.e., construction phase, assessment 1, refinement phase, and assessment 2). Trained_1 and Trained_2 were the solution rates of the training items in assessment 1 and assessment 2, respectively. In our design, Trained_1 was used as a measure of learning for the knowledge construction phase and as a measure of prior knowledge for the knowledge refinement phase. Transfer_1 and Transfer_2, in contrast, depended on the specific type of transfer item. Specifically, in the models predicting performance on unambiguous, ambiguous, or intra-range items, Transfer_1 and Transfer_2 were the solution rates for the respective type of transfer item in either assessment 1 or assessment 2. To make a structurally comparable model for extra-range items, which were only presented in assessment 2, we used the solution rates on intra-range items as Transfer_1. Each of these variables was regressed on all the preceding variables, with the exception of Trained_1 \rightarrow Transfer_2 to avoid redundancy of the Trained_2 \rightarrow Transfer_2 path.

Second, we computed the effects of reasoning ability, WMC, and SSR at each stage of the relational category learning task. We used the sum score of Set II in the RAPM as a measure of reasoning ability. For the three WMC tasks, we used partial credit scoring to compute the score for each individual task (i.e., first averaging over items per trial, then averaging over trials). As an overall measure of WMC, we intended to create a latent variable using the solution rates in the memory updating, sentence span, and spatial short-term memory tasks. As shown in Table 5.1, we found surprisingly low correlations between the spatial short-term memory and the memory updating tasks ($r = .23$), as well as between the spatial short-term memory and the sentence span tasks ($r = .10$), which makes the extraction of a latent variable both problematic and questionable. We therefore decided to remove the spatial short-term memory task from all analyses and to base our WMC factor on the memory updating and sentence span tasks.\footnote{In the original paper introducing the WMC battery (Lewandowsky et al., 2010), the correlations were slightly larger, ranging from .32 to .48 for the spatial short-term memory and the memory updating} However, when
freely estimating the loading of the sentence span task and using memory updating as the marker variable, the theta matrix contained negative residual variances for the memory updating task in two models. Therefore, we fixed the loading of the sentence span task in all models to 0.39, reflecting the averaged loadings found in the non-problematic models (i.e., $0.389 + 0.391 \div 2$). This approach had the additional advantage that the WMC measurement model was identical in all models. Regarding SSR, we created parcels reflecting strategy use during the learning phases and during the assessments by averaging over the two respective scores. These parcels had high reliabilities (see Table 5.1) and were used as indicators for the SSR latent variable, with both loadings being fixed to 1 (see Figure 5.4).

Third, since we were not only interested in direct effects, we also investigated indirect (i.e., mediated) effects. Therefore, for each significant direct path originating in RAPM, WMC, and SSR, we investigated all the indirect effects on learning and transfer performance in exploratory analyses. For example, assuming a direct effect of SSR → Transfer_1, we would investigate the following indirect paths: SSR → Transfer_1 → Trained_2, SSR → Transfer_1 → Transfer_2 and SSR → Transfer_1 → Trained_2 → Transfer_2 (see Figure 5.4). We tested the significance of the indirect effects (as well as the direct effects) with a bootstrapping procedure using 10000 random draws (MacKinnon, Lockwood, & Williams, 2004). Bias-corrected 95% confidence intervals were provided and an effect was judged as significant if zero was not included in the interval.

Finally, as a sanity check, we recomputed both the repeated measures ANOVAs for the learning phases and the structural equation models with an additional predictor coding for the two stimulus sets used in the learning paradigm. Stimulus set was not significantly related to the outcome in any model (all $p$'s > 0.05), which is why we did not further consider this predictor in any of the analyses reported in this paper.

---

tasks, and from .20 to .28 for the spatial short-term memory and the sentence span tasks. The shortened version of spatial short-term memory task used in the present study might be responsible for this discrepancy. However, this is unlikely given that the mean, standard error, and Cronbach’s alpha in our sample and the values reported in Lewandowsky et al. (2010) were very similar (see Table 5.1). Further, we still obtained a relatively high correlation between the spatial short-term memory task and RAPM ($r = 36$; see Table 5.1).
References


References


References


6 General Discussion

6.1 Integrative Summary of the Main Findings

The present thesis aimed at illuminating the association between reasoning ability, working memory capacity (WMC), and conceptual learning. Several open questions were raised at the beginning of this thesis and three empirical studies have been conducted to address them (see section 2.5). The studies have the distinctive feature that they are crossing boundaries between research areas such as educational psychology, cognitive/experimental psychology, and cognitive neuroscience with respect to the investigated issues and the applied methods. In the remainder of this thesis, I want to first provide an integrative summary of the main findings, mention strengths and weaknesses of the research, give an outlook and formulate implications, and end with an overall conclusion.

The goal of the first study was to better characterize the neurocognitive basis of WMC and reasoning ability. Although every human is inevitably (and more often than not annoyingly so) aware that one cannot perform arbitrarily complex tasks, i.e., that one’s working memory is limited, the reason for that is still under scrutiny. There are theoretical accounts assuming that some individuals are better at attention control, which helps them to stay on task and maintain information, especially when exposed to distracting and interfering elements (Engle, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004). In contrast, storage capacity accounts stress that working memory is limited by the number of elements that can be simultaneously represented (Cowan, 2001; Cowan et al., 2005; Luck & Vogel, 1997; Miller, 1956). Recently, another emphasis has been put on the fact that some individuals might be more adept at retrieving only the relevant information from long-term memory and disregard concurrently activated but irrelevant information (Unsworth & Engle, 2007). The findings of empirical studies conducted in recent years about the nature of WMC and why it relates to reasoning ability have been quite heterogeneous. For that reason, we applied neu-
6.1 Integrative Summary of the Main Findings

roimaging, specifically electroencephalography (EEG), during working memory tasks involving these different mechanisms to gain an additional perspective. Among other more specific findings, we found that brain activity with respect to all those mechanisms could differentiate between participants with higher and lower reasoning ability, while behavioral performance was uninformative in this regard. In particular, reasoning ability was related to the flexible adaption of brain activity in the alpha band, which is assumed to be an indicator of the functional inhibition of brain areas, to the specific task requirements. Overall, this was interpreted in line with recent accounts arguing that not a single but multiple mechanisms underlie the association between working memory and reasoning ability (e.g., Conway & Kovacs, 2015; Shipstead, Lindsey, Marshall, & Engle, 2014).

The second study was concerned with characterizing the role of these working memory mechanisms for conceptual learning in school. Since the sample was recruited from a classroom study that investigated the growth in conceptual knowledge about Newtonian mechanics resulting from 18 lessons of instruction in Swiss higher secondary school, we could draw on an ecologically valid measure of conceptual learning. Although the ability to stay on task and inhibit irrelevant information (attention control), the maintenance of information (storage capacity) and controlled retrieval from long-term memory are all theoretically expected to be important for conceptual learning, we found that only the latter was uniquely predictive of learning gains. This is well in line with the essential role of prior knowledge for learning, in particular for the acquisition of conceptual knowledge (e.g., Carey, 2000; Stern, 2015; Tricot & Sweller, 2013). For example, in domains such as physics where students bring everyday preconceptions to the classrooms that are incompatible with scientific concepts (e.g., everyday explanations about force and inertia), successful learning requires that both are activated and explicitly compared (Carey, 2000; Clement, 1982). Thus, the ability to narrowly constrain memory retrieval to the relevant information (and disregard irrelevant or even disagreeing information) might have helped the learners to develop a coherent conceptual knowledge base during the instruction. We also analyzed brain activity with EEG during working memory performance, but in contrast to the first study did not find that neural measures (neither oscillatory activity nor functional connectivity) provided information over and above the behavioral measures.

Finally, the third study explicitly examined the influence of prior knowledge, to-
6.1 Integrative Summary of the Main Findings

together with cognitive abilities, on conceptual learning. Because both are more often than not correlated in real life, arguably due to interactions during development, we created a novel laboratory-based category learning paradigm that with its high experimental control allowed us to better investigate their specific effects during conceptual learning.

The learning paradigm was comprised of two learning phases: In the first phase, the participants learned to diagnose two artificial diseases that were defined by an XOR-relation between the values of two medical tests. In the second phase, this (now prior) knowledge was required to accurately learn when treatment was required and when it was not required. We found that reasoning ability predicted conceptual learning in the first phase, but its effect on learning in the second phase was only indirect (i.e., mediated over its effect on prior knowledge). At the same time, prior knowledge was the strongest predictor with respect to learning in the second phase, but working memory provided an additional positive effect. This is an experimental corroboration of correlational findings showing that cognitive ability (in this case, reasoning ability) becomes less predictive when prior knowledge is included (Lohman, 1999; Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2013; Primi, Ferrão, & Almeida, 2010; Weinert, Helmke, & Schneider, 1990; Woodrow, 1946). However, these findings also suggest that this is only half the truth: Reasoning ability still had a substantial effect on future learning, but it was mediated over its initial effect on prior knowledge. Furthermore, somewhat reminiscent of the findings from the second study, working memory came into play in the second learning phase, i.e., when prior knowledge had to be retrieved and integrated with novel knowledge.

In addition to the interplay between cognitive abilities and prior knowledge, we also assessed learning strategies in the third study. After each learning phase, we presented participants with specific transfer items that allowed us to determine whether they tended to follow a rule-based strategy (i.e., trying to abstract the rule governing the disease classification) or an exemplar-based strategy (i.e., learning the test values and corresponding diseases by heart). As expected, the rule-based strategy outperformed the exemplar-based strategy particularly on transfer items that shared no superficial similarity with the training items. This is particularly relevant with respect to today’s science education, since one of its majors aims is to convey knowledge that can be transferred in a flexible manner to many different examples and contexts (De Corte, 2003). However, it also shows that - unsurprisingly - factors beyond cognitive abilities and prior knowl-
6.2 Strengths and Limitations

The major strength of the present studies is that we investigated issues and applied methods from multiple research areas, including educational psychology, cognitive psychology, and cognitive neuroscience. The first study used neuroimaging, in particular EEG, to answer questions that have raised lively interest in cognitive psychology, specifically why WMC is so predictive of higher cognitive functions such as intelligence and reasoning ability. By "zooming in" on and visualizing the specific cognitive processes (Stern & Schneider, 2010), we were able to compare the most prominent psychological theories from a neural perspective. This constitutes another instance demonstrating that neuroscientific research well-grounded in psychological theory can provide valuable new insights. In the second study, to somewhat mitigate the limited ecological validity of typical neuroscience experiments, we related both behavioral and neural measures collected in the laboratory to a measure of conceptual learning in real classrooms. In the third study, we drew on literature from both cognitive and educational psychology in an attempt to better integrate both fields with respect to category and conceptual learning. Specifically, we used the distinctive experimental control that category learning paradigms performed in the laboratory provide to better understand the role of cognitive abilities and prior knowledge in conceptual learning.

In contrast, as usual in empirical research, there are also some limitations. With respect to the first two studies, we used different working memory tasks representative for the respective mechanisms rather than experimentally manipulating these mechanisms within a single coherent task. Although this allowed us to directly draw on previous findings, it might have introduced unwanted "method variance". Furthermore, a general problem of neuroscientific research, which also concerns the present studies, is the oftentimes too limited sample size (Button et al., 2013). Replications with larger, ideally more heterogeneous samples, would help to better assess the robustness and generalizability of the present findings. Directly related to this issue, the participants of the presented studies were either Gymnasium (which prepares for university in Switzerland; first and second study) or university students (third study). Although we still found pre-
6.3 Outlook and Implications

The present thesis provides important insights into the interplay between reasoning ability, WMC, and prior knowledge during conceptual learning. However, while the investigated constructs are clearly relevant for education and understanding their relation to conceptual learning is important, and while the research questions were also inspired by educational research (e.g., the effect of prior knowledge and cognitive abilities on learning), the present findings do not by themselves have practical implications as to how instruction in classrooms should be designed. There are, however, implications for future research, some of which will be discussed in the following.

Regarding the first study of this thesis, further studies following a similar approach should consider applying functional magnetic resonance imaging (fMRI) to extend the measurement of brain activity (see also section 6.2). The combination of the high temporal resolution of EEG with the high spatial resolution of fMRI would make it possible to better localize the functional differences in brain activity to specific brain networks: For example, differences in attention control could potentially be mapped to connections...
6.3 Outlook and Implications

between the prefrontal cortex, anterior cingulate cortex, and other subcortical structures; differences in maintenance might be mapped to connections between prefrontal and parietal areas; and differences in retrieval mechanisms might be mapped to connections between prefrontal cortex and hippocampal areas (e.g., Conway, Getz, Macnamara, & Engel de Abreu, 2011; Jonides et al., 2008). Modern techniques to analyze functional connectivity such as Granger causality and Dynamic Causal Modeling might be particularly fruitful in this regard (Friston, Moran, & Seth, 2013). The extension of the methodological repertoire and the refinement of approaches will step-by-step help to improve our understanding of the neural processes underlying interindividual differences in cognitive abilities.

Furthermore, I believe that category learning paradigms similar to that used in the third study have the potential to become an important tool to investigate and inform educationally relevant issues. They make the mechanisms of the acquisition of conceptual knowledge more traceable as compared to classroom studies and can readily be adapted to examine all sorts of research questions. For example, in our case, we did not provide any instructional guidance from "outside" to shape the learning of the participants. The items were presented one after another and the order of items from each category (i.e., disease) was random. However, an approach that has been successfully applied to help learning in both laboratory and classrooms is prompting learners to compare two exemplars that are shown side by side (e.g., Alfieri, Nokes-Malach, & Schunn, 2013; Gick & Holyoak, 1983; Kurtz, Boukrina, & Gentner, 2013). The comparison process is thought to promote the extraction of the common structure underlying the exemplars, which in turn facilitates transfer to exemplars with a similar structure but different superficial features (Gentner, 1983, 2010). Incorporating the comparison of exemplars into the learning paradigm from the third study would not only make it possible to see how comparison influences initial knowledge construction in the first learning phase, but also how knowledge refinement is affected in the second learning phase. The advantage of comparing exemplars might become particularly apparent when incremental learning is investigated. Another potentially beneficial instructional method is modifying the order of presentation so that exemplars from different categories are alternately presented. This interleaving of exemplars, compared to a blocked presentation of exemplars (or problems) from the same category, is thought to improve learning by highlighting the differences that are particularly relevant for classification (for a review see
6.4 Conclusion

Rohrer, 2012). Furthermore, category learning paradigms could also be easily accommodated to include other successful instructional methods such as self-explanations or meta-cognitive questions (Beeth, 1998; Chi, Leeuw, Chiu, & LaVancher, 1994). Overall, the evaluation of such methods in a well-controlled setting will help to increase our understanding of how and when they improve learning.

The investigation of conceptual learning within the scope of category learning paradigms might also help to bring neuroscientific and educational research closer together. The application of neuroimaging methods generally requires that the tasks are relatively simple and well-structured so that multiple trials can be averaged to obtain an acceptable signal-to-noise-ratio. Accordingly, much research in cognitive neuroscience with a link to education has applied relatively basic (oftentimes mathematical) tasks such as number comparisons (Ansari, 2008) or mental arithmetic (Zamarian, Ischebeck, & Delazer, 2009). However, although important from a basic research point of view, the investigated issues are often not directly informed by education nor are the findings directly relevant for educational practice. The fact that category learning paradigms are structurally relatively simple, but as outlined above, also highly adaptable to examine educationally relevant issues, makes them an ideal candidate for an educationally inspired neuroscience. For example, such a paradigm would allow for the investigation of how the effect of prior knowledge is reflected in brain activity, how its effect differs from that of cognitive abilities, whether the effects of different instructional methods can be differentiated in terms of involved neural structures and whether one approach might be more promising than another in this regard. The examination of such interdisciplinary questions will, in the end, benefit both cognitive neuroscience and educational research.

6.4 Conclusion

At the beginning of this thesis, I stated that understanding what makes humans different is a necessary prerequisite towards understanding the human mind. In an attempt to shed some light on these differences, this thesis presented three studies to examine the interrelationship between reasoning ability, WMC, and conceptual learning with the help of methods from multiple research areas. Although novel insights about when and why these constructs are related were gained and the potential of such an approach was
demonstrated, many questions remain unanswered. However, this is really the essence of science, which is why I hope that the presented work will inspire future research.

References


References


References


147
References


Appendix
Probandeninformation (Version vom 03.11.2014)

Neuronale Grundlagen von Intelligenz und Lernfähigkeit

a) Zielsetzung
Das Ziel der Studie ist es herauszufinden, welche kognitiven Fähigkeiten (z.B. sich Informationen merken oder diese wieder abrufen) am stärksten mit intellektuellen Fähigkeiten und Lern- und Schulerfolg zusammenhängen. Hierzu wird nicht nur die Leistung (Genauigkeit, Schnelligkeit) in kognitiven Aufgaben, sondern auch die Hirnaktivität mit einem Elektroenzephalogramm (EEG) erfasst.

b) Ablauf
Als Erstes werden Sie gebeten von zu Hause aus einen anonymisierten Online-Fragebogen auszufüllen. Diese beinhaltet Fragen zu Ihrer Person und Ihrer Händigkeit. Zusätzlich kommen Sie für eine ca. 3-stündige Sitzung an die ETH Zürich (Gebäude IFW, Etage B, Haldeneggsteig 4, 8092 Zürich). Bitte bringen Sie zu diesem Termin Ihr aktuellstes Schulzeugnis (oder eine Kopie davon) mit.


In der ersten Aufgabe sehen Sie abwechselnd einfache arithmetische Gleichungen (z.B. 3+4=7), die Sie als richtig oder falsch bewerten sollen und Buchstaben, die Sie sich merken sollen. Nach vier bis neun Gleichungen bzw. Buchstaben müssen Sie die Buchstaben wiedergeben.


Im dritten Test wird Ihnen mehrmals ein Bild mit mehreren farbigen Quadraten gezeigt. Ihre Aufgabe ist es, sich die Farben für eine kurze Zeit zu merken und bei einem zweiten gezeigten Bild zu entscheiden, ob es gleich oder anders ist als das erste Bild.


Im fünften Test merken Sie sich mehrmals die Reihenfolge von vier Positionen in einem Gitter und sehen dann vier einfache Gleichungen (z.B. 3+4=7), die Sie als richtig oder falsch bewerten sollen. Danach müssen Sie die eben gelernten Positionen im Gitter auswählen.

Nach der Computer-Testung werden Ihnen die Elektroden und die EEG-Kappe abgenommen, und Sie können sich die Haare waschen, um die aufgetragene Leitpaste (siehe d) zu entfernen. Daraufhin werden Sie für ca. 30 min schriftliche Tests zur Messung der geistigen Leistungsfähigkeit lösen. Abschliessend werden Sie ausführlich über die Ziele der Studie informiert und erhalten die Aufwandsentschädigung (siehe f).

c) Bedingungen für die Studienteilnahme
Eine Teilnahme an dieser Studie ist möglich bei Zutreffen der folgenden Kriterien:

- schriftliche Einwilligung
- wenn jünger als 18 Jahre: schriftliche Einwilligung der Eltern
- keine Farbenblindheit
- keine psychische Erkrankungen und Lernstörungen (z.B. Lese- oder Rechenschwäche)
- keine neurologische Erkrankungen
- keine regelmässige Einnahme von psychotropen Medikamenten oder Drogen

Bitte informieren Sie den Untersucher falls einer oder mehrere Ausschlussgründe vorliegen oder Sie eine Frage haben. Welcher Ausschlussgrund vorliegt, muss nicht angegeben werden.

d) Vor- und Nachteile für Probanden/Innen / Risiken
Die Studienteilnahme ist mit einer finanziellen Entschädigung und Rückmeldung über die individuelle Leistung verbunden (siehe f). Darüber hinaus können durch die Studie wertvolle Einsichten darüber gewonnen werden, was intellektuelle Fähigkeiten sowie Lern- und Schulerfolg ausmachen. Sie erhalten ausserdem Einblick in aktuelle neurowissenschaftliche Forschung zum menschlichen Lernen und Denken. Ein direkter pädagogischer Nutzen kann jedoch nicht erwartet werden kann.


e) Finanzierung
Das Projekt wird durch Forschungsmittel der ETH Zürich finanziert. Weder den Teilnehmenden noch deren Krankenkassen werden studienbedingte Kosten (z.B. der EEG-Testung) in Rechnung gestellt.

f) Entschädigung

g) Rücktrittsrecht
Sowohl die Studienteilnehmer als auch deren Eltern können die Mitwirkung an der Studie jederzeit beenden, ohne Angabe von Gründen und ohne dass ihnen dadurch Nachteile entstehen.

h) Datenschutz

i) Versicherungsschutz
Allfällige Gesundheitsschäden, die in direktem Zusammenhang mit der Studie entstehen und auf nachweisliches Verschulden der ETH Zürich zurückzuführen sind, sind durch die Betriebs-Haftpflichtversicherung der ETH Zürich (Police Nr. 100.001 der Schweizerischen Mobiliar Versicherungsgesellschaft) gedeckt. Darüber hinaus liegt die Unfall-/Krankenversicherung (z.B. für die Hin- und Rückreise) in der Verantwortung der Probandin/des Probanden.

j) Kontaktperson
Die Studie wird durchgeführt von M. Sc. Bruno Rütsche an der Abteilung für Lehr- und Lernforschung, Institut für Verhaltenswissenschaften der ETH Zürich, Universitätstrasse 41, 8092 Zürich. Er steht Ihnen für Fragen gerne zur Verfügung:

E-Mail: bruno.ruetsche@ifv.gess.ethz.ch
Tel.: 044 632 50 53
Einverständniserklärung (Version vom 03.11.2014)

⇒ Bitte lesen Sie dieses Formular sorgfältig durch.
⇒ Bitte fragen Sie den/die Untersucher/in oder Ihre Kontaktperson, wenn Sie etwas nicht verstehen oder etwas wissen möchten.

Titel der Studie: Neuronale Grundlagen von Intelligenz und Lernfähigkeit

Durchführungsort der Studie: ETH Zürich, Professur für Lehr- und Lernforschung
Gebäude IFW, Raum B31
Haldeneggsteig 4
8092 Zürich

Untersucher/in (Name und Vorname): .................................................................

Proband/in (Name und Vorname): .................................................................

⇒ Die Studie wird durchgeführt unter der Leitung von M. Sc. Bruno Rütsche an der Abteilung für Lehr- und Lernforschung, Institut für Verhaltenswissenschaften der ETH Zürich, Universitätstrasse 41, 8092 Zürich. Er steht Ihnen für Fragen gerne zur Verfügung (bruno.ruetsche@ifv.gess.ethz.ch, Tel.: 044 632 50 53).
⇒ Ich nehme an dieser Studie freiwillig teil und kann jederzeit ohne Angabe von Gründen meine Zustimmung zur Teilnahme widerrufen, ohne dass mir deswegen Nachteile entstehen.
⇒ Ich wurde mündlich und schriftlich über die Ziele, den Ablauf der Studie, über die zu erwartenden Wirkungen, über mögliche Vor- und Nachteile sowie über eventuelle Risiken informiert.
⇒ Ich hatte genügend Zeit, um meine Entscheidung zu treffen.
⇒ Ich bestätige mit meiner Unterschrift, dass ich die im Informationsblatt genannten Bedingungen für die Studienteilnahme erfülle.
⇒ Ich bin darüber informiert, dass die allgemeine Haftpflichtversicherung der ETH Zürich (Police Nr. 100.001 der Schweizerischen Mobiliar Versicherungsgesellschaft) nur Gesundheitsschäden deckt, die in direktem Zusammenhang mit der Studie entstehen und auf nachweisliches Verschulden der ETH Zürich zurückzuführen sind. Darüber hinaus liegt die Unfall-/Krankenversicherung (z.B. für die Hin- und Rückreise) in meiner Verantwortung.
⇒ Ich bin einverstanden, dass die zuständigen Untersuchenden und/oder Mitglieder der Ethikkommission zu Prüf- und Kontrollzwecken meine Originaldaten einsehen dürfen, jedoch unter strikter Einhaltung der Vertraulichkeit.
⇒ Ich bin mir bewusst, dass während der Studie die in der Probanden-Information genannten Anforderungen und Einschränkungen einzuhalten sind. Im Interesse meiner Gesundheit kann mich die untersuchende Person auch ohne gegenseitiges Einverständnis von der Studie ausschliessen. Zudem orientiere ich die untersuchende Person über das Vorliegen von möglichen Ausschlussgründen der Untersuchungsteilnahme.
⇒ Zudem orientiere ich die untersuchende Person über die gleichzeitige Behandlung bei einem Arzt sowie über die Einnahme von Medikamenten (vom Arzt/von der Ärztin verordnete oder selbständig gekauften).
⇒ Ergeben sich während der Studie Zufallsergebnisse, die zur Diagnose, Behandlung oder Verhinderung bestehender oder künftig drohender Krankheiten führen können, möchte ich 
☐ darüber aufgeklärt werden. ☐ nicht darüber aufgeklärt werden.

Ort, Datum ............................. Unterschrift Proband/in .............................

Ort, Datum ............................. Unterschrift Erziehungsberechtigte/r .............................
(falls Proband/in jünger als 18)

Ort, Datum ............................. Unterschrift Untersucher/in .............................
Probandeninformation (Version vom 14.09.2015)
Komplexe Phänomene lernen

a) Zielsetzung

b) Ablauf
Sie kommen für zwei ca. 1.5-stündige Sitzungen ins Decision Science Laboratory der ETH Zürich (ETH Zentrum, Gebäude IFW, Etage A, Haldeneggsteig 4, 8092 Zürich). In der ersten Sitzung werden Sie eine Lernaufgabe mit zwei Teilen bearbeiten. Im ersten Teil werden Sie über mehrere Durchgänge hinweg lernen, wie zwei Krankheiten diagnostiziert werden. Der zweite Teil baut auf dem ersten auf und Sie lernen, wann bei diesen Krankheiten eine Behandlung notwendig ist und wann nicht. Anschliessend lösen Sie noch eine kurze Aufgabe in der mehrmals zwei Bilder zu vergleichen sind.

In der zweiten Sitzung werden Sie zunächst mehrere Gedächtnis-Aufgaben am Computer machen. In der ersten Aufgabe sehen Sie abwechselnd einfache Sätze, die Sie als richtig oder falsch bewerten sollen, und Buchstaben, die Sie sich merken und anschliessend wiedergeben sollen. In der nächsten Aufgabe merken Sie sich mehrere Zahlen, wobei sich diese Zahlen mehrmals durch die Anwendung einfacher arithmetischer Operationen ändern, und geben die Zahlen anschliessend wieder. In der dritten Aufgabe merken Sie sich mehrere Positionen in einem 10x10 Gitter und geben diese dann wieder. Hiernach werden Sie am Computer einen Test zur Messung der geistigen Leistungsfähigkeit machen. Sowohl innerhalb wie auch zwischen den einzelnen Aufgaben haben Sie die Möglichkeit, Pausen zu machen.

Am Ende der Untersuchung werden Sie ausführlich über die Ziele und Untersuchungsbedingungen der Studie informiert und erhalten die Aufwandsentschädigung (siehe Punkt f).

c) Bedingungen für die Studienteilnahme
Eine Teilnahme an dieser Studie ist möglich bei Zutreffen der folgenden Kriterien:

- mindestens 18 Jahre alt
- Student/-in
- kein Studium in Medizin oder Pharmazie
- keine psychischen Erkrankungen und Lernstörungen (z.B. Lese- oder Rechenschwäche)
• keine neurologische Erkrankungen
• keine regelmässige Einnahme von psychotropen Medikamenten oder Drogen

Bitte informieren Sie den Untersucher falls einer oder mehrere Ausschlussgründe vorliegen oder Sie eine Frage haben. Welcher Ausschlussgrund vorliegt, muss nicht angegeben werden.

d) Vor- und Nachteile für Probanden/Innen / Risiken
Die Studienteilnahme ist mit einer finanziellen Entschädigung verbunden (siehe f). Darüber hinaus können durch die Studie wertvolle Einsichten darüber gewonnen werden wie komplexe Informationen gelernt werden und welche Faktoren das Lernen beeinflussen.

e) Finanzierung
Das Projekt wird durch Forschungsmittel der ETH Zürich finanziert. Weder den Teilnehmenden noch deren Krankenkassen werden studienbedingte Kosten in Rechnung gestellt.

f) Entschädigung
Sie erhalten als Entschädigung für die Studienteilnahme 80 CHF in bar.

g) Rücktrittsrecht
Sie können Ihre Mitwirkung an der Studie jederzeit beenden, ohne Angabe von Gründen und ohne dass Ihnen dadurch Nachteile entstehen.

h) Datenschutz
In dieser Studie werden persönliche Daten von Ihnen erhoben. Alle Daten werden reversibel anonymisiert und für 10 Jahre aufbewahrt. Sie sind nur Fachleuten zur wissenschaftlichen Auswertung zugänglich. Die zuständige Ethikkommission kann zu Kontrollzwecken Einsicht in die Originaldaten nehmen. Während der ganzen Studie und bei den erwähnten Kontrollen wird die Vertraulichkeit strikt gewahrt. Ihr Name wird in keiner Weise in Rapporten oder Publikationen, die aus der Studie hervorgehen, veröffentlicht.

i) Versicherungsschutz
Allfällige Gesundheitsschäden, die in direktem Zusammenhang mit der Studie entstehen und auf nachweisliches Verschulden der ETH Zürich zurückzuführen sind, sind durch die Betriebs-Haftpflichtversicherung der ETH Zürich (Police Nr. 100.001 der Schweizerischen Mobiliar Versicherungsgesellschaft) gedeckt. Darüber hinaus liegt die Unfall-/Krankenversicherung (z. B. für die Hin- und Rückreise) in der Verantwortung der Probandin/des Probanden.

j) Kontaktperson
Die Studie wird M. Sc. Bruno Rütsche unter der Leitung von Dr. sc. Lennart Schalk (Professur für Lehr- und Lernforschung, Institut für Verhaltenswissenschaften, ETH Zürich) durchgeführt. Sie stehen Ihnen für Fragen gerne zur Verfügung:

E-Mail: bruno.ruetsche@ifv.gess.ethz.ch
Tel.: 044 632 50 53
Einverständniserklärung (Version vom 14.09.2015)

⇒ Bitte lesen Sie dieses Formular sorgfältig durch.
⇒ Bitte fragen Sie den/die Untersucher/in oder Ihre Kontaktperson, wenn Sie etwas nicht verstehen oder etwas wissen möchten.

Titel der Studie: Komplexe Phänomene lernen

Durchführungsort der Studie: ETH Zürich, Professur für Lehr- und Lernforschung
Gebäude IFW, IFW A25-29
Haldeneggsteig 4
8092 Zürich

Untersucher/in (Name und Vorname): .................................................................

Proband/in (Name und Vorname): .................................................................

⇒ Die Studie wird von M. Sc. Bruno Rütsche unter der Leitung von Dr. Sc. Lennart Schalk (Professur für Lehr- und Lernforschung, Institut für Verhaltenswissenschaften, ETH Zürich) durchgeführt. Sie stehen Ihnen für Fragen gerne zur Verfügung: bruno.ruetsche@ifv.gess.ethz.ch, Tel.: 044 632 50 53
⇒ Ich nehme an dieser Studie freiwillig teil und kann jederzeit ohne Angabe von Gründen meine Zustimmung zur Teilnahme widerrufen, ohne dass mir deswegen Nachteile entstehen.
⇒ Ich wurde mündlich und schriftlich über die Ziele, den Ablauf der Studie, über die zu erwartenden Wirkungen, über mögliche Vor- und Nachteile sowie über eventuelle Risiken informiert.
⇒ Ich hatte genügend Zeit, um meine Entscheidung zu treffen.
⇒ Ich bestätige mit meiner Unterschrift, dass ich die im Informationsblatt genannten Bedingungen für die Studienteilnahme erfülle.
⇒ Ich bin darüber informiert, dass die allgemeine Haftpflichtversicherung der ETH Zürich (Police Nr. 100.001 der Schweizerischen Mobilar Versicherungsgesellschaft) nur Gesundheitsschäden deckt, die in direktem Zusammenhang mit der Studie entstehen und auf nachweisliches Verschulden der ETH Zürich zurückzuführen sind. Darüber hinaus liegt die Unfall-/Krankenversicherung (z.B. für die Hin- und Rückreise) in meiner Verantwortung.
⇒ Ich bin einverstanden, dass die zuständigen Untersuchenden und/oder Mitglieder der Ethikkommission zu Prüf- und Kontrollzwecken meine Originaldaten einsehen dürfen, jedoch unter strikter Einhaltung der Vertraulichkeit.
⇒ Ich bin mir bewusst, dass während der Studie die in der Probandeninformation genannten Anforderungen und Einschränkungen einzuhalten sind. Im Interesse meiner Gesundheit kann mich die untersuchende Person auch ohne gegenseitiges Einverständnis von der Studie ausschliessen. Zudem orientiere ich die untersuchende Person über das Vorliegen von möglichen Ausschlussgründen der Untersuchungsteilnahme.

Ort, Datum ....................... Unterschrift Proband/in ....................

Ort, Datum ....................... Unterschrift Untersucher/in .................

155
Task Instructions: Stroop

- In dieser Aufgabe werden Ihnen mehrfach die Wörter «ROT», «GRÜN» oder «BLAU» in einer bestimmten Druckfarbe gezeigt.
- Die Druckfarbe kann entweder zum Wort passen (d.h. ROT, GRÜN, BLAU) oder nicht (d.h. ROT, ROT, GRÜN, GRÜN, BLAU, BLAU).
- Ihre Aufgabe ist es, auf die Druckfarbe zu achten und die entsprechend eingefärbte Taste zu drücken.
- Falls Sie zu langsam antworten, geht es automatisch weiter mit der nächsten Aufgabe.

¹The code of all experimental paradigms is available upon request. Please contact Bruno Rütsche (bruno.ruetsche@ifv.gess.ethz.ch).
**Task Instructions: Stroop**

- Verwenden Sie den Zeige-, Mittel- und Ringfinger und lassen Sie diese während des gesamten Experiments auf den Tasten liegen.
- Vermeiden Sie während der gesamten Aufgabe Augen- und Kopfbewegungen. Schauen Sie stets in die Mitte des Bildschirms (Fixationspunkt).
- Es sind mehrere Pausen zur Erholung eingebaut.
- Es ist wichtig, dass Sie während des gesamten Experiments konzentriert bleiben. Bitte antworten Sie so schnell und korrekt wie möglich.
Task Instructions: Visual Arrays

In diesem Test werden Ihnen für eine kurze Zeit 2, 4 oder 6 farbige Quadrate an zufälligen Positionen gezeigt.

Ihre Aufgabe ist es, sich das gezeigte Bild zu merken.

Danach wird Ihnen ein zweites Bild gezeigt, in dem ein einzelnes Quadrat umkreist ist. Das zweite Bild ist entweder genau gleich wie das erste oder unterscheidet sich in der Farbe des umkreisten Quadrats. Die Position und Anzahl der Quadrate sowie die Farbe der nicht umkreisten Farben ändern sich nicht zwischen beiden Bildern.

Entscheiden Sie so schnell wie möglich, ob das zweite Bild gleich oder anders ist als das erste. Dafür benutzen Sie bitte die mit "richtig" oder "falsch" beschrifteten Tasten.
Task Instructions: Visual Arrays

- Verwenden Sie den Zeige- und Mittelfinger und lassen Sie diese während des gesamten Experiments auf den Tasten liegen.

- Vermeiden Sie während der gesamten Aufgabe Augen- und Kopfbewegungen. Schauen Sie stets in die Mitte des Bildschirms (Fixationspunkt).

- Es sind mehrere Pausen zur Erholung eingebaut.

- Es ist wichtig, dass Sie während des gesamten Experiments konzentriert bleiben. Bitte antworten Sie so schnell und korrekt wie möglich.
In diesem Test gibt es drei Phasen: Merk-, Rechen- und Abrufphase.

In der Merkphase werden Ihnen nacheinander vier schwarze Quadrate an unterschiedlichen Positionen innerhalb eines Gitters mit 16 Feldern (4 x 4) gezeigt.

Merken Sie sich bitte die Positionen in genau der Reihenfolge, in der sie präsentiert werden.
Task Instructions: Secondary Memory

- In der darauffolgenden Rechenphase sehen Sie nacheinander vier einfache Gleichungen.
- Sie sollen jedes Mal bewerten, ob die Gleichung richtig oder falsch ist. Zum Beispiel, «4 + 3 = 7» ist richtig, aber «10 − 1 = 8» ist falsch.
- Jede Gleichung erscheint für nur drei Sekunden auf dem Bildschirm. Sie müssen Ihre Entscheidung in dieser Zeit abgeben. Dafür benutzen Sie bitte die linke Maustaste für richtig und die rechte Maustaste für falsch. Es ist wichtig, dass Sie mindestens zu 85% korrekte Antworten geben, also vermeiden Sie bitte während Ihrer Antwortgabe Fehler.

- Die darauffolgende Abrufphase wird durch ein «?» unter dem Gitter signalisiert.
- Sie sollen sofort nach dem Erscheinen des «?» die vorhin gelernte Reihenfolge von Positionen abrufen.
- Sobald Sie alle Positionen, an die Sie sich erinnern konnten, abgerufen haben, drücken Sie die linke Maustaste.
- Nach dem Drücken erscheint der bisher unsichtbare Mauszeiger in der Mitte des Gitters.
- Wählen Sie dann mit der linken Maustaste nacheinander die Positionen aus.
- Wenn Sie sich an eine Position nicht erinnern konnten, lassen Sie diese nicht aus, sondern raten Sie.
Verwenden Sie den Zeige- und Ringfinger und lassen Sie diese während des gesamten Experiments auf der Maus liegen.

Vermeiden Sie während der gesamten Aufgabe Augen- und Kopfbewegungen. Schauen Sie stets in die Mitte des Bildschirms (Fixationspunkt).

Es sind mehrere Pausen zur Erholung eingebaut.

Es ist wichtig, dass Sie während des gesamten Experiments konzentriert bleiben. Bitte antworten Sie so schnell und korrekt wie möglich.
Task Instructions: Updating

1. In diesem Test gibt es zwei Phasen: Merk- und Abrufphase.
2. In der Merkphase werden Ihnen zuerst nacheinander drei farbige Quadrate (Rot, Grün, Blau) an unterschiedlichen Positionen innerhalb eines Gitters mit 16 Feldern (4 x 4) gezeigt.
3. Danach erscheint vier- bis sechsmal ein Pfeil in der Mitte des Gitters. Dieser hat die Farbe eines der Quadrate (d.h., ←, ↙, ↗), und zeigt in eine bestimmte Richtung (←, ↑, →, ↘).
4. Dies bedeutet, dass Sie das Quadrat, das der Farbe des Pfeils entspricht, einen Schritt in die angezeigte Richtung bewegen sollen.
5. Ihre Aufgabe ist es, sich die letzte Position jeder Farbe zu merken.
Task Instructions: Updating

- Die darauffolgende **Abrufphase** wird durch ein «?» unter dem Gitter signalisiert.
- In dieser Phase müssen Sie nacheinander die zuvor gemerkten Farb-Positions-Kombinationen abrufen.
- Wählen Sie mit der linken Maustaste die Position aus.
- Wenn Sie sich an eine Position nicht erinnern, lassen Sie diesen nicht aus, sondern raten Sie.
Task Instructions: Updating

- Verwenden Sie den Zeige- und Ringfinger und lassen Sie diese während des gesamten Experiments auf der Maus liegen.
- Vermeiden Sie während der gesamten Aufgabe Augen- und Kopfbewegungen. Schauen Sie stets in die Mitte des Bildschirms (Fixationspunkt).
- Es sind mehrere Pausen zur Erholung eingebaut.
- Es ist wichtig, dass Sie während des gesamten Experiments konzentriert bleiben. Bitte antworten Sie so schnell und korrekt wie möglich.
Lernaufgabe

Diagnose von Krankheiten

Mit den Pfeiltasten können Sie vor- und zurückgehen.

Bitte lesen Sie die folgenden Erklärungen sorgfältig durch. Bei Fragen melden Sie sich bitte bei der Versuchsleitung.
Task Instructions: Category Learning - Phase 1

Patientenunterlagen

- Stellen Sie sich vor, Sie reisen als Ersatz für einen verstorbenen Arzt in eine abgelegene Urwaldregion.
- Der verstorbene Arzt hat entdeckt, dass die Bewohner dieser Region an einer von zwei Krankheiten leiden: Midosis oder Buriosis.
- Ihr Vorgänger hat Patientenunterlagen hinterlassen, in denen er für jeden Patienten die Ergebnisse von zwei medizinischen Tests und die daraus diagnostizierte Krankheit festgehalten hat.
- Sie versuchen die Krankheiten anhand der Patientenunterlagen zu lernen.

Patientenunterlagen

- Zum Lernen nutzen Sie ein Computerprogramm.
- Dieses Programm zeigt Ihnen die Ergebnisse zweier medizinischer Tests (Test A und Test B).
- Sie müssen dann bei jedem Patienten feststellen, welche der zwei Krankheiten (Midosis oder Buriosis) vorliegt. Klicken Sie dazu mit der Maus auf die entsprechende Schaltfläche.
- Danach erhalten Sie von dem Programm direkt eine Rückmeldung, ob Sie sich «richtig» oder «falsch» entschieden haben (siehe Bild auf der nächsten Seite).
Task Instructions: Category Learning - Phase 1

Neue Patienten

- Nachdem Sie gelernt haben, kommen neue Patienten in Ihre Praxis.
- Sie führen mit diesen Patienten die gleichen medizinischen Tests wie bei den bereits bekannten Patienten aus den Unterlagen durch.
- Sie müssen ebenfalls bei jedem Patienten feststellen, welche der zwei Krankheiten vorliegt. Klicken Sie dazu mit der Maus auf die entsprechende Schaltfläche.
- Da Sie für die neuen Patienten nicht auf die Unterlagen des verstorbenen Arztes zurückgreifen können, erhalten Sie keine Rückmeldung, ob Ihre Antwort «richtig» oder «falsch» war.
Es werden Ihnen immer mehrere Patienten nacheinander präsentiert. Nach jeweils 16 Patienten gibt es die Möglichkeit, eine kurze Pause zu machen.

Im Laufe dieser Aufgabe werden Sie 224 Patienten diagnostizieren.

Bitte verhalten Sie sich über die ganze Aufgabe hinweg ruhig, um die anderen Teilnehmer nicht zu stören.

Es ist wichtig, dass Sie während des gesamten Experiments konzentriert bleiben.

Zusammenfassung:

Ihre Aufgabe ist es, die Krankheiten anhand der Patientenunterlagen zu lernen und danach neue Patienten zu diagnostizieren.
Lernaufgabe
Behandlung der Krankheiten

Mit den Pfeiltasten können Sie vor- und zurückgehen.

Bitte lesen Sie die folgenden Erklärungen sorgfältig durch.
Bei Fragen melden Sie sich bitte bei der Versuchsleitung.
**Patientenunterlagen**

- Bei einer weiteren Durchsuchung des Archivs des verstorbenen Arztes haben Sie zusätzliche Patientenunterlagen entdeckt.
- In diesen Unterlagen beschreibt er, dass bei seinen Patienten nicht immer eine Behandlung notwendig war.
- Der Arzt hat für jeden seiner Patienten auch den Behandlungsbedarf festgehalten.
- Sie versuchen daher anhand der Patientenunterlagen zu lernen, wann eine Behandlung notwendig ist und wann nicht.

---

**Patientenunterlagen**

- Zum Lernen nutzen Sie ein Computerprogramm.
- Dieses Programm zeigt Ihnen erneut die Ergebnisse der zwei medizinischen Tests (Test A und Test B).
- Sie müssen nun bei jedem Patienten feststellen, welche der zwei Krankheiten (Miosis oder Buriosis) vorliegt und ob eine Behandlung notwendig ist oder nicht. Klicken Sie dazu mit der Maus auf die entsprechende Schaltfläche.
- Danach erhalten Sie von dem Programm direkt eine Rückmeldung, ob Sie sich «richtig» oder «falsch» entschieden haben (siehe Bild auf der nächsten Seite).
Der Name der Krankheit ist angegeben.

Test A 203
Test B 329

Burkosis
Mikrosis
Kernische
Zellular
Zellscheibe
Zellscheibe
Schaltfläche

Neue Patienten

• Nachdem Sie gelernt haben, kommen neue Patienten in Ihre Praxis.
• Sie führen mit diesen Patienten die gleichen medizinischen Tests wie bei den bereits bekannten Patienten aus den Unterlagen durch.
• Sie müssen ebenfalls bei jedem Patienten feststellen, welche der zwei Krankheiten vorliegt und ob eine Behandlung notwendig ist oder nicht. Klicken Sie dazu mit der Maus auf die entsprechende Schaltfläche.
• Da Sie für die neuen Patienten nicht auf die Unterlagen des verstorbenen Arztes zurückgreifen können, gibt es keine Rückmeldung darüber, ob Ihre Antwort «richtig» oder «falsch» war.
Task Instructions: Category Learning - Phase 2

- Es werden Ihnen immer mehrere Patienten nacheinander präsentiert. Nach jeweils 16 Patienten gibt es die Möglichkeit, eine kurze Pause zu machen.
- Im Laufe dieser Aufgabe werden Sie 240 Patienten diagnostizieren.
- Bitte verhalten Sie sich über die ganze Aufgabe hinweg ruhig, um die anderen Teilnehmer nicht zu stören.
- Es ist wichtig, dass Sie während des gesamten Experiments konzentriert bleiben.

Zusammenfassung:

Ihre Aufgabe ist es zu lernen, wann bei den Krankheiten eine Behandlung notwendig ist und wann nicht.