# From Labels to Functions

## How Working Memory Capacity facilitates processing of Matrix Reasoning items with Multiple Rules

**Dissertation**

Zur Erlangung des akademischen Grades eines

*Doktors der Philosophie*

der Fakultät HW, Bereich für Empirische Humanwissenschaften der Universität des Saarlandes

vorgelegt von

**Florian Krieger (geb. Domnick)**

aus Berlin

Saarbrücken, 2018

| | |
|---|---|
| Dekan: | Univ.-Prof. Dr. Cornelius König, Universität des Saarlandes |
| 1. Berichterstatter: | Univ.-Prof. Dr. Hubert Zimmer, Universität des Saarlandes |
| 2. Berichterstatterin: | Univ.-Prof. Dr. Gisa Aschersleben, Universität des Saarlandes |
| Tag der Disputation: | 09. April 2018 |

# Acknowledgement

An dieser Stelle möchte ich vielen Personen danken, die mich während meiner Zeit als Doktorand begleitet und beraten haben und es mir ermöglicht haben, diese Arbeit fertigzustellen. Ich danke Alexander Kirmße, Matthias Stadler und Ashley Johnson für hilfreiche Kommentare zu einer früheren Fassung dieser Arbeit. Speziell danke ich Nicolas Becker, der mich mit der Faszination zur Thematik der Matrizentests bereits während meines Bachelor-Studiengangs angesteckt hat und mich seitdem kompetent unterstützt und begleitet. Zudem war es für mich eine sehr wertvolle Erfahrung im Graduiertenkolleg „Adaptive Minds" promovieren zu dürfen und ich möchte mich hierbei bei allen Kollegen für die tollen Jahre, bei Axel Mecklinger als Sprecher des Kollegs, und bei Frank Spinath für das Vertrauen in mich bedanken. Ein besonderer Dank geht an meinen Betreuer Hubert Zimmer, der mir zum einen immer freie Hand gelassen hat, mich zu entwickeln, mich zum anderen aber immer mit großer Geduld und Kompetenz unterstützt hat, wenn ich Hilfe brauchte – teils in mehrstündigen Sitzungen ☺.

Neben meinen Kollegen gibt es eine Reihe an Personen, die mir auf den letzten drei Jahren sehr geholfen haben. Diese Danksagung bietet allerdings nicht den passenden Raum, um den Beitrag aller Personen in gebührender Weise zu würdigen. Ich möchte aber dennoch vier Personen kurz hervorheben: meine Eltern, Antonia und vor allem Andrea.

Part of this work is also included in one published article and one article in preparation for submission (see citation below). Please note that the first article was published under my “maiden name” *Domnick*. To warrant a smooth reading, the respective passages are not marked in the text. In addition, following the practice of these articles, I employ “we” instead of “I” for the entire work.

Domnick, F., Zimmer, H. D., Becker, N., & Spinath, F. M. (2017). Is the Correlation between Storage Capacity and Matrix Reasoning Driven by the Storage of Partial Solutions? A Pilot Study of an Experimental Approach. *Journal of Intelligence*, 5.

Krieger, F., Zimmer, H.D., Greiff S., Spinath, F.M, & Becker N. (in prep). Selective Encoding Demands in Matrix Reasoning: The Role of Working Memory Capacity.

# Table of contents


***Acknowledgement .......... 3***

***Table of contents .......... 5***

***List of Figures .......... 8***

***List of Tables .......... 9***

***List of Abbreviations .......... 10***

***Abstract .......... 11***

***Deutsche Zusammenfassung .......... 12***

***A Introduction .......... 1***

***B Matrix Reasoning .......... 5***

**1 Significance of Matrix Reasoning in gF .......... 5**

**2 Invention of Matrix Reasoning .......... 8**

**3 Item Difficulty in Matrix Reasoning .......... 9**

**4 Processing in Matrix Reasoning .......... 11**

4.1 Models on Geometric Analogies .......... 12

4.2 Matrix-Reasoning Processing .......... 13

4.3 Single-Rule Processing .......... 14

4.4 Multiple-Rule Processing .......... 16

**5 Interim Conclusion 1 .......... 22**

***C Working Memory Capacity .......... 24***

**1 The Concept of Working Memory .......... 24**

**2 What is Working Memory Capacity? .......... 26**

2.1 Storage Capacity .......... 27

2.2 Controlled Attention .......... 28

2.3 WMC in a Nutshell .......... 35

**3 WMC and gF .......... 36**

3.1 Storage + Processing and gF ........................................................................ 37
3.2 Filtering and gF ........................................................................ 38
3.3 Storage Capacity and gF ........................................................................ 39
3.4 Implications on Matrix Reasoning ........................................................................ 40
***D Matrix Reasoning and WMC ........................................................................ 43***
**1 Multiple Rules and WMC ........................................................................ 43**
**2 Shortcomings in Studies on WMC and Matrix Reasoning ........................................................................ 44**
**3 Interim Conclusion 2 ........................................................................ 45**
***E The Current Study ........................................................................ 46***
***F Study 1: Partial Solutions ........................................................................ 50***
**1 Introduction ........................................................................ 50**
**2 Methods ........................................................................ 51**
2.1 Participants and Design ........................................................................ 51
2.2 Test Methods ........................................................................ 51
2.3 Data Analysis ........................................................................ 55
**3 Results ........................................................................ 56**
**4 Discussion ........................................................................ 58**
***G Study 2: Filtering (Preliminary Study for Study 3) ........................................................................ 60***
**1 Introduction ........................................................................ 60**
**2 Methods ........................................................................ 61**
2.1 Participants and Design ........................................................................ 61
2.2 Test Methods ........................................................................ 61
2.3 Data Analysis ........................................................................ 64
**3 Results ........................................................................ 65**
3.1 Control of variances ........................................................................ 66
3.2 Filter costs ........................................................................ 66
**4 Discussion ........................................................................ 67**
***H Study 3: Selective Encoding and Goal Management ........................................................................ 69***
**1 Study 3A ........................................................................ 70**
1.1 Method ........................................................................ 71

1.2 Results ..... 76
1.3 Discussion ..... 78

**2 Study 3B ..... 79**

2.1 Methods ..... 80
2.2 Results ..... 85
2.3 Discussion ..... 90

***I General Discussion ..... 91***

**1 Overview ..... 91**

**2 Implications on Matrix Reasoning ..... 91**

2.1 Partial Solutions ..... 91
2.2 Selective Encoding ..... 96
2.3 Goal Management ..... 100

**3 Implications on WMC ..... 102**

3.1 Storage Capacity and Storage + Processing ..... 102
3.2 Filtering ..... 106

**4 A Revised Process-Model of Matrix Reasoning ..... 107**

**5 Conclusion ..... 110**

***J Appendix ..... 111***

***K References ..... 113***

# List of *Figures*

Figure 1: Example of a matrix-reasoning item ........ 1
Figure 2: Matrix reasoning item with one rule and four rules ........ 2
Figure 3. Simplified CHC-model of intelligence ........ 7
Figure 5: Illustration of the intersection rule ........ 10
Figure 4: Illustration of the addition rule ........ 10
Figure 6. Illustration of item difficulties ........ 11
Figure 7: Example of a geometric analogy ........ 12
Figure 8. Process model of matrix reasoning ........ 13
Figure 9: Illustrate of an item with one rule ........ 14
Figure 10. Illustration of an item with multiple rules ........ 16
Figure 11. Illustration of the embedded process model ........ 25
Figure 12: Illustration of a change detection task with color change. ........ 27
Figure 13: Procedure of the Operation Span ........ 31
Figure 14. Storage and processing task and filtering tasks ........ 32
Figure 15: Procedure of the filter task ........ 33
Figure 16. Different perspectives on WMC and its measurement. ........ 35
Figure 17. Design of the current work. ........ 46
Figure 18. Procedure of the change detection task for Study 1. ........ 52
Figure 19. Illustration of the two versions of matrices test for Study 1 ........ 54
Figure 20. Results of SEM in Study 1 ........ 58
Figure 21. Result of path analyses in Study 2 ........ 66
Figure 22. Item examples in Study 3 ........ 73
Figure 23. Results of Study 3A. ........ 77
Figure 24. Results of Study 3B ........ 87
Figure 25. Results of SEM in Study 3B ........ 88
Figure 26. Illustration of an item applied in the present study and items with more difficult correspondence finding ........ 99
Figure 27. Revised process-model of matrix reasoning ........ 108

# List of Tables

*Table 1.* Descriptives and correlations of Study 1 .... 57

*Table 2*. Descriptives of all tasks in Study 2 .... 65

Table 3. Inter-correlations of all tasks in Study 2 .... 65

*Table 4.* Descriptives of Study 3A .... 76

*Table 5*. Descriptives of Study 3B .... 86

# List of Abbreviations

$\eta^2$ ....... Eta square as effect size for a one-way ANOVA
ANOVA ....... analysis of variance
APM ....... Advanced Progressive Matrices
cf. ....... compare
e.g. ....... for example
DESIGMA ....... Design a Matrix test
df ....... degrees of freedom
EXT ....... externalized condition of Study 1
*g* ....... intelligence
gF ....... fluid intelligence
K ....... K-index of storage capacity
mm ....... millimeter
ms ....... milliseconds
NonEXT ....... non-externalized condition of Study 1
S+P ....... storage plus processing
*p* ....... probability of significance given that null hypothesis is true
*r* ....... Pearson Product-Moment Correlation Coefficient
ROI ....... region of interest
SD ....... standard deviation
*t* ....... test statistic from Student's *t*-distribution
WMC ....... working memory capacity

# Abstract

The question why some individuals are more intelligent than others is one of the most important questions of the last 100 years in psychology. This study set out to investigate why individuals are better in matrix reasoning as one of the most prominent proxies of intelligence. One well-replicated finding is that matrix-reasoning items with multiple rules are harder to solve than items with a single rule. Notably, it is assumed that the individual working memory capacity (WMC) plays a crucial role in the processing of items with multiple rules. However, it is still an ongoing question *why* WMC is facilitating the processing of these items. In this work, we investigated possible roles of WMC in matrix-reasoning items with multiple rules. In doing so, we experimentally manipulated certain processes in matrix reasoning which are suggested in the literature to be more demanded in items with multiple rules. In addition, we observed the impact of WMC on these processes from a functional perspective. That is to say, we defined WMC not as an overall resource, but based on the WMC-literature, we examined which aspect of WMC could be required for the respective processes in matrix reasoning. The **first study** investigated whether storing partial solutions is required in matrix-reasoning and whether individual *storage capacity* as one aspect of WMC facilitates the storing of partial solutions. The **second study** can be regarded as a preliminary study for the third study, which quantified the influence of *filtering* as a further aspect of WMC on matrix-reasoning. The **third study** investigated whether selective encoding demands are present in multiple-rule items by means of both behavioral and eye movement analyses. We also observed whether individual *filtering* ability facilitates selective encoding in matrix reasoning. In addition, we observed whether goal management demands are present in multiple-rule items and whether individual *storage and processing* as another aspect of WMC is related to goal management. Results of all studies revealed that neither storing partial solutions nor goal management were required in multiple rule items, nor that these demands were associated with the aspects of WMC assessed in the respective studies. In contrast, results indicate that higher difficulties in multiple-rule items were mainly due to higher demands on selective encoding and more importantly, *filtering* facilitated processing of items with these demands. The results of the present study entail important implications for both matrix-reasoning processing and intelligence but also for our understanding of the involvement of WMC in intelligence.

# Deutsche Zusammenfassung

Die Frage, warum manche Menschen intelligenter sind als andere, ist eine der wichtigsten Fragen der letzten 100 Jahre in der Psychologie. In dieser Studie wurde untersucht, warum einige Personen in figuralen Matrizentests – als einer der prominentesten Verfahren zu Erfassung von Intelligenz – besser sind als andere. Ein gut replizierter Befund ist, dass figuralen Matrizentest-Aufgaben mit mehreren Regeln schwieriger zu lösen sind als Aufgaben mit einer einzigen Regel. Insbesondere wird davon ausgegangen, dass die individuelle Arbeitsspeicherkapazität (WMC) eine entscheidende Rolle bei der Verarbeitung von Aufgaben mit mehreren Regeln spielt. Es ist jedoch immer noch ungeklärt, warum WMC die Bearbeitung dieser Aufgaben erleichtert. Deshalb untersuchten wir in dieser Arbeit mögliche Einflüsse von WMC in figuralen Matrizentest mit mehreren Regeln. Hierbei manipulierten wir experimentell bestimmte Prozesse in figuralen Matrizentests, die in der Literatur als wichtige Prozess diskutiert werden, die bei Aufgaben mit mehreren Regeln stärker beansprucht zu sein scheinen. Darüber hinaus beobachteten wir den funktionalen Einfluss von WMC auf diese Prozesse. Das heißt, wir haben WMC nicht als Gesamtressource definiert, sondern auf der Grundlage der Literatur untersucht, welcher Aspekt von WMC für die jeweiligen Prozesse in figuralen Matrizentests benötigt werden könnte. Die erste Studie untersuchte, ob die Speicherung von Teillösungen in figuralen Matrizentest erforderlich ist und ob die individuelle Speicherkapazität, als Teilaspekt von WMC, die Speicherung von Teillösungen erleichtert. Die zweite Studie kann als Vorstudie für die dritte Studie betrachtet werden, die den Einfluss der Filterfähigkeit als weiteren Aspekt von WMC auf figural Matrizentests quantifizierte. Die dritte Studie untersuchte anhand von Verhaltens- und Augenbewegungsanalysen, ob selektive Enkodierungsanforderungen in Aufgaben mit mehreren Regeln vorhanden sind. Wir beobachteten zudem, ob individuelle Filterfähigkeiten das selektive Enkodieren in figuralen Matrizentests erleichtert. Darüber hinaus beobachteten wir, ob Anforderungen an das Zielmanagement in Aufgaben mit mehreren Regeln vorhanden sind und ob die Fähigkeit Inhalte im Arbeitsgedächtnis während der Bearbeitung einer kompetitiven Zeitaufgabe zu speichern mit dem Zielmanagement zusammenhängt. Die Ergebnisse der Studien zeigen, dass weder die Speicherung von Teillösungen noch das Zielmanagement in Aufgaben mit mehreren Regeln erforderlich war, noch, dass diese Anforderungen mit den jeweiligen Aspekten des WMC, die in den jeweiligen Studien erhoben wurde, zusammenhing. Im Gegensatz deuten die Ergebnisse darauf hin, dass höhere Schwierigkeiten bei Aufgaben mit mehreren Regeln hauptsächlich auf höhere Anforderungen an die selektive Enkodierung zurückzuführen waren, und was noch wichtiger ist, die Filterfähigkeit das Lösen dieser Aufgaben erleichterte. Die Ergebnisse der vorliegenden Studie beinhalten wichtige Implikationen sowohl für die Verarbeitung von figuralen Matrizentests als auch für Intelligenz im Allgemeinen, aber auch für unser Verständnis über die Beteiligung von WMC an Intelligenz.

"Therefore, investigating WMC, and its relationship with intelligence, is psychology's best hope to date to understand intelligence." (Oberauer, Schulze, Wilhelm, & Süß, 2005)

# A Introduction

*What is the nature of human intelligence, and why are some individuals more intelligent than others?*

These questions have provided the basis for a large deal of research in psychology over the last decades, especially because intelligence is a strong predictor for vital aspects of life such as school grades (Roth et al., 2015), job performance (Hunter, 1986) or even health (Gottfredson & Deary, 2004). In various established models of intelligence, fluid intelligence (gF) is considered as a critical aspect of intelligence (e.g. Carroll, 1993; Horn & Cattell, 1966a), which can be described as the ability to adapt to novel situations, independent of prior knowledge or experience (Cattell, 1963). One of the most prominent tests to assess gF is the matrix-reasoning test. The current study is set out to investigate the processing of matrix-reasoning to gain further insights into the nature of gF.

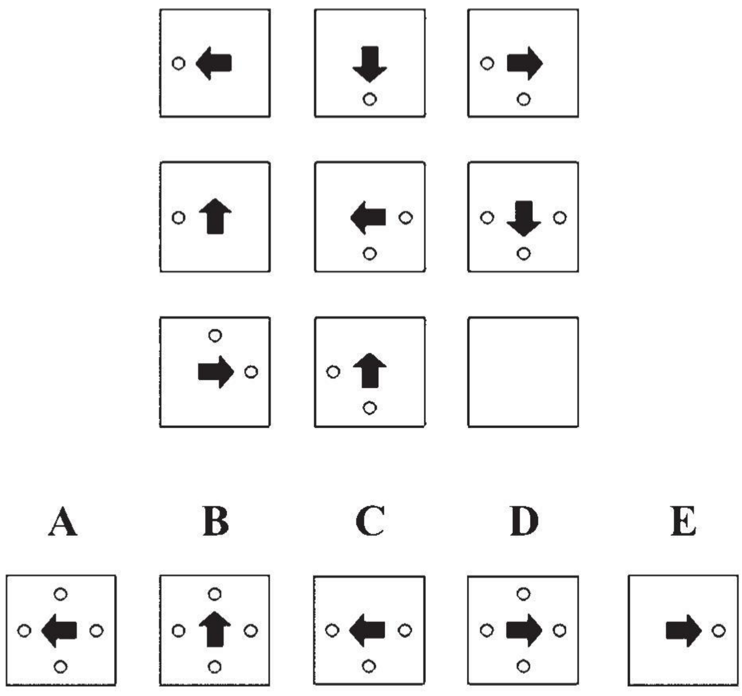



*Figure 1*: Example of a matrix-reasoning item (Becker, Preckel, Karbach, Raffel, & Spinath, 2014, p. 2)

A matrix reasoning task commonly consists of a 3×3 matrix filled with several visual elements that follow underlying design rules (see *Figure 1*). Within the matrix, the lower right field (solution field) is usually left empty and must be filled according to the applied rules. In *Figure 1*, the arrow is rotating row-wise from cell to cell by 90 degrees. Additionally, the circles of the first two cells are summed up in the third cell. When the rules are successfully applied to the last row, it can be inferred that answer A is the correct solution that fits into the solution field.

Due to its ease of administration and fast evaluation, matrix reasoning is preferentially used to assess gF, and therefore, matrix reasoning is also included in well-established intelligence assessment batteries, such as the Wechsler Adult Intelligence Scale (WAIS–IV, Wechsler, 2008).

Another advantage of matrix-reasoning tests is that items can be created, which are highly demanding and therefore, can only be solved by a few individuals. This is critical since a test of gF is only useful when it discriminates between good and bad performer. It is well-replicated that one of the most important determinants of item difficulty is the number of applied rules in a matrix (Becker, Schmitz, Göritz, & Spinath, 2016; Carpenter, Just, & Shell, 1990; Green & Kluever, 1992; Meo, Roberts, & Marucci, 2007; Primi, 2002; Vigneau, Caissie, & Bors, 2006; Vodegel Matzen, van der Molen, & Dudink, 1994).

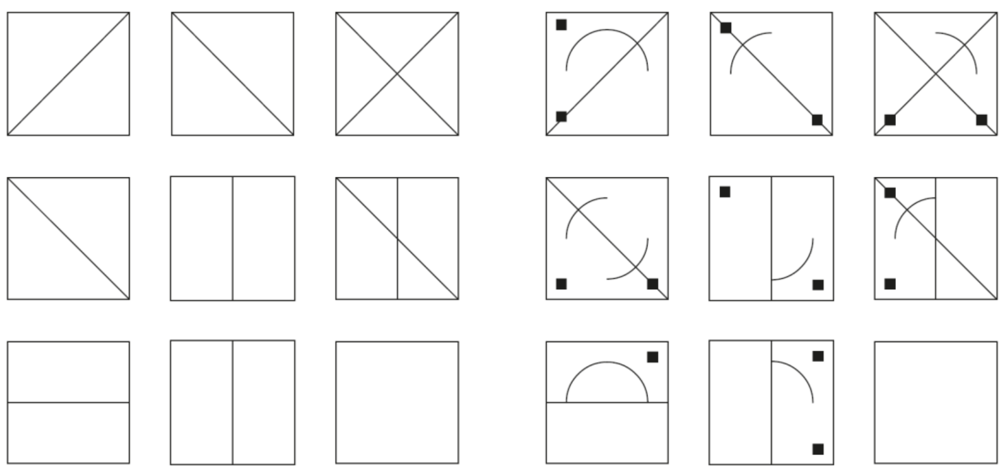


*Figure 2*: Matrix reasoning item with one rule (left) and four rules (right; Items adapted from the DESIGMA; Becker et al., 2014)

To demonstrate this effect, two matrices are contrasted with each other in *Figure 2*; the left matrix contains one rule and the right matrix four rules. When solving these two matrices, it appears that the right matrix is harder to solve than the left matrix as more information has to be taken into consideration.

When we revert to the initial question why some individuals are more intelligent than others, we can ask in terms of matrix reasoning: Why do some individuals outperform others in a matrix-reasoning task, especially when multiple rules are applied? One assumption that seems auspicious in this regard is that individual working memory capacity (WMC) limits the processing of elements in matrix-reasoning (e.g. Carpenter et al., 1990). WMC is described as the number of distinct pieces of information that can be held active for further processing (Cowan, 2001), and also how effective this information can be encoded and maintained (Engle, 2002). Consequently, respondents with low WMC can only attend to few figural elements or lack in the effectiveness of storing these elements and therefore, are failing in finding the correct solution. The critical role of WMC in matrix-reasoning is supported by substantial correlations between WMC and matrix-reasoning (Harrison, Shipstead, & Engle, 2015; Jarosz & Wiley, 2012; Loesche, Wiley, & Hasselhorn, 2015; Salthouse, 1993; Unsworth & Engle, 2005), which is also in line with findings from latent-variable approaches that demonstrate the significant role of WMC in gF (Ackerman, Beier, & Boyle, 2005; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle et al., 1999; Unsworth, Fukuda, Awh, & Vogel, 2014).

Although these studies demonstrate the critical role of WMC in both gF and matrix reasoning, they cannot clearly resolve *why* WMC should facilitate processing in matrix reasoning. In other words, these studies cannot solve the issue why items with multiple rules are harder to solve than items with one rule.

The current project is set out to investigate the *why* with regard to two important core aspects. On the one hand, process models of matrix reasoning suggest that WMC has an influence on different processes in solving. Therefore, it was one goal of this work to isolate these processes and to consider the influence of WMC on these processes separately. Here, we wanted to focus on the processes of storage of partial solutions (e.g, Mulholland, Pellegrino, & Glaser, 1980), selective encoding (e.g., Primi, 2002), and goal management (e.g., Carpenter et al., 1990).

On the other hand, several studies demonstrated that the relation between WMC and gF is driven by different sources underlying WMC (e.g., Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth, Fukuda, Awh, & Vogel, 2014). That is to say, WMC is a result of different *aspects* or *processes*, which enable a successful active maintenance of information. Hence, we considered WMC not unitary diagnostic under the *label* "WMC", which is covered by one single task or task set. In contrast, based on the WMC-literature, we examined which aspect of WMC could be required for the respective processes in matrix reasoning.

In the next chapters, we will first give an overview of matrix reasoning with the aim to describe the solving process in more detail. Subsequently, we will introduce a theoretical basis of WMC, explain which aspects of WMC were relevant for this work, and how they are related to gF. Finally, we describe the relationships between the aspects of WMC and the matrix-reasoning processes considered in this study.

# B Matrix Reasoning

In this chapter, we will first briefly describe how matrix reasoning evolved and is embedded in well-established intelligence theories to illustrate the significance of matrix-reasoning. Second, we will outline what specifically determines item difficulty to describe item characteristics which distinguish good performers from bad performers. At this point, we will primarily focus on the number of rules in a matrix as the main predictor of item difficulty. Finally, we will describe which processes are required when solving an item with multiple rules to gain an insight as to why items with multiple rules are harder to solve than items with a single rule.

## 1 Significance of Matrix Reasoning in gF

When reviewing well-established models of intelligence, it is salient that all models regard the solving of *novel problems* as an essential mechanism of intelligence, which requires the ability "to adapt effectively to the environment […], to engage in various forms of reasoning, to overcome obstacles by taking thought" (Neisser et al., 1996, p. 77).

Solving novel problems as an important part of intelligence was already highlighted in early intelligence models, such as the theory of *general intelligence* by Spearman (1904). Spearman found that all ability and school performance tests shared a substantial amount of variance and were loading on a higher-order general-factor *g.* In addition, Spearman (1904) described that each task also loaded a second, task-specific factor, which did not share variance with *g*. Thus, if one considers four performance tests, all would load on *g*, and each test would also have a task-specific variance proportion, which did not share variance with *g*. Spearman considered *g* as a mental "energy" which is involved in every mental task and which became over the decades *the* synonym for intelligence. Notably, in a later work, Spearman (1927) argued that *g* is involved in solving novel problems. To this end, he defined two laws: the *eduction of relations* and the *eduction of correlates*. The first aspect was described as the detection of a relation between two or more elements. The second was described as “any idea together with a relation” (Spearman, 1927, p. 166), which indicated the detection of an underlying rule of the elements. The word eduction can be derived from the Latin word *educare*, which means "to draw out" and thus can be described as a process of making sense out of given material. More importantly, Spearman (1927) described that eduction refers to solving novel problems, and thus to the relation of elements that were not previously known.

Identifying a relationship between elements in novel problems was further explored in later intelligence models. For example, by means of factor analyses, Thurstone (1938) was able to identify *reasoning* as one important primary ability for intelligence besides other primary abilities. As eduction, reasoning also describes the detection of a relation between given, novel elements. Interestingly, Thurstone and Thurstone (1941) could demonstrate that a factor similar to reasoning was closely related to a superordinate *g* factor. They termed this factor *induction*, which, again, describes the detection of a rule in a given, novel material. Both the fact that this description was very close to the description of *eduction* as well as the high factor loading on *g* supported Spearman's (1927) initial assumption that detecting relations between elements in novel problems is an important part of intelligence. Hence, all terms (eduction, reasoning, induction) describe the same ability, which is the ability to detect an underlying rule in a novel problem.

The most widely used term for this ability today is fluid intelligence (gF), which was introduced by Cattel and Horn (Cattell, 1963; Horn, 1968; Horn & Cattell, 1966). As eduction, reasoning or induction, gF describes solving and adapting to *novel* situations without relying on previous learning experience. For instance, finding the underlying rule of the *letter series* (a, z, y, a, z, ?) to infer the missing letter. Importantly, to solve this task the respondent does not necessarily needs to have a concept of the letters or needs to know the alphabet. Instead, the respondent only needs to induce the underlying rule by finding regularities in the sequence. This can also be illustrated by the fact that the rules of letters could also be represented with other material such as different pictures or numbers.

In the model by Cattel and Horn (1966), gF stays in contrast to abilities, which rely on previous learned knowledge, which they termed as crystallized intelligence (gC). An example task set for gC is verbal reasoning like the task *"...is to water like eating is to..."*. The respondent has to select the correct answer, which is presented along several distractors (A. Travelling-Driving, B. Foot-Enemy, C. Drinking-Bread, D. Girl-Industry, and E. Drinking-Enemy; example taken from the Differential Aptitude Test; e.g., Martínez & Colom, 2009). In this example, C would be the correct answer. Since the respondent can only solve the task when he or she knows the concept or the vocabulary, these tasks rely on acquired knowledge and, therefore, are not independent of cultural influences.

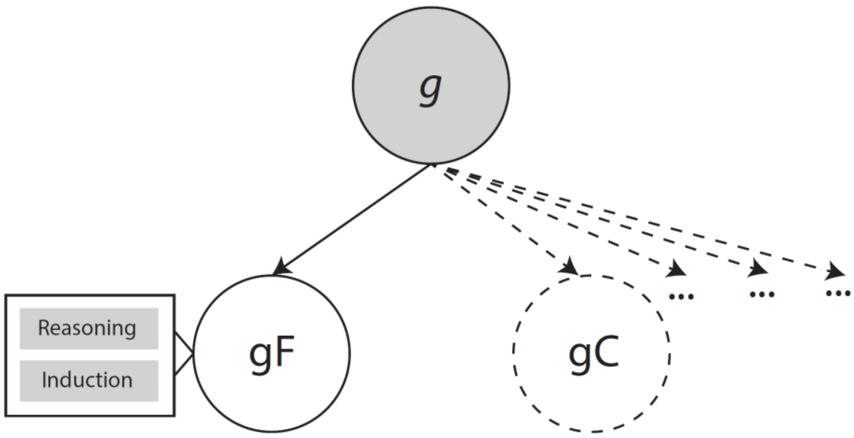



*Figure 3*. Simplified CHC-model of intelligence (McGrew, 2009). Abilities of interest for the current study are highlighted.

The role of gF and gC in intelligence was also emphasized in other models, for example in the Three-stratum theory by Carroll (1993), and current theories of intelligence also consider gF and gC important abilities of intelligence. For instance, in the Cattell-Horn-Carroll theory (CHC-theory; McGrew, 2009), gF and gC are essential abilities, which load on a super-ordinate *g* factor, which describes general intelligence as in Spearman's (1904) theory. It is important to note that also other abilities are considered in this model (as also in Three-stratum theory; Carroll, 1993). Notably, gF is described in the CHC model by the narrow abilities *induction* and *reasoning*, which again emphasizes the role of detection rules of elements in novel problems (cf., McGrew, 2009).

For the current work, we define intelligence according to the CHC-model, in which *g* is on a superordinate level, which covers all abilities that are connected with intelligence (see *Figure 3*). In particular, we use gF as an umbrella term for the detection of rules in novel problems such as eduction, reasoning or induction.

# 2 Invention of Matrix Reasoning

To test gF, several tests were developed, such as number or letter series. The most prominent test, however, is matrix reasoning. One of the first known rationals to design such a test was described by Spearman (1927; as cited in Jensen, 1998). Spearman's intention was that tests should examine the eduction of relations and correlates. Hence, as described above, the test should require the detection of an underlying rule of a novel problem independent of pre-learned knowledge.

John C. Raven, a student of Spearman, captured this idea and developed a test that contained several matrix-reasoning problems. Starting with the Standard Progressive Matrices (SPM; Raven, 1938), he later introduced the more difficult Advanced Progressive Matrices (APM; Raven, 1940). Although other matrix-reasoning tests were developed in the last decades (e.g., Bochumer Matrizentest, Hossiep & Hasella, 2010; Wiener Matrizentest, Formann & Piswanger, 1979), the APM can still be seen as a gold standard to assess gF or *g* (e.g., Rost, 2009). In fact, Jensen résumés that "I have yet to see a factor analysis of any diverse collection of tests that includes Raven's Matrices in which the Raven's largest loading was found on any factor other than *g*" (Jensen, 1998, p. 38).

The study by Marshalek, Lohman, and Snow (1983) supports the crucial role of the APM in intelligence research. Based on multidimensional scaling of a large quantity of ability tests, the authors formulated the (revised) Radex model. The more a task is placed in the center of this model, the more "complex" is the task. That means, that these tasks are more correlated with *g* and are requiring the induction of rules and abstract problem solving. Importantly, the APM were found to be placed in the center of the model indicating a high association with *g*, and more precisely with gF. Importantly, the tasks of the surrounding are also requiring the induction of rules in novel problems, which underlines the affinity of the APM with these demands.

Taken together, matrix reasoning is considered as a paramount tool to assess *g*, or more specifically gF, as it requires the induction of underlying rules by inferring the correct solution without previous knowledge (Eysenck, 1998; Jensen, 1998; Neisser et al., 1996; Rost, 2009). Therefore, it is no surprise that researcher attempt to uncover what determines the difficulty of matrix reasoning items in order to derive indications for the nature of *g* or gF. The aim of these studies is to identify characteristics in matrix-reasoning items that are only solved by a few participants, and therefore, discriminates between good and bad performer. The next section will address the main characteristics that determine item difficulty.

# 3 Item Difficulty in Matrix Reasoning

There are two main characteristics that influence the difficulty in matrix reasoning (cf., Green & Kluever, 1992). The first characteristic is the *type of rule*. In a matrix-reasoning item, different rules are applied. For example, *Figure 4* shows an addition rule and *Figure 5* an intersection rule.

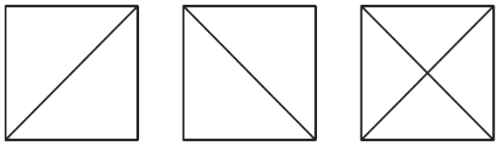


*Figure 4*: Illustration of the addition rule (adapted from the DESIGMA; Becker et al., 2014)

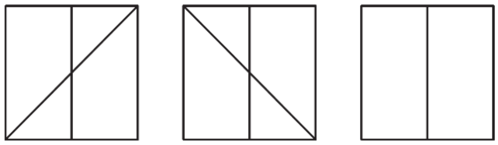


*Figure 5*: Illustration of the intersection rule (adapted from the DESIGMA; Becker et al., 2014)

In the addition rule, the elements of the first two cells are added in the third cell, whereas in the intersection rule, only those elements appear in the third cell that occur in both the first and second cell. Several studies have shown that the type of the applied rules in a matrix vary in their difficulty (Carpenter et al., 1990; Embretson, 1998; Green & Kluever, 1992). This means that some rules are solved by almost all individuals and some rules by very few. For example, it could be shown that an item with the addition rule is solved more often than an item with the intersection rule (Becker et al., 2014).

Although the type of rule is mainly influencing the item difficulty, we did not focus on this determinant in the present study. However, we counterbalanced the type of rule in the studies of the present work to reduce this effect as a possible disturbance.

More important for the present work was the number applied rules in a matrix. In fact, a large pool of studies could demonstrate that the number of applied rules in a matrix is the main determinant of item difficulty (Becker, Schmitz, Göritz et al., 2016; Carpenter et al., 1990; Embretson, 1998; Green & Kluever, 1992; Meo et al., 2007; Primi, 2002; Vigneau et al., 2006; Vodegel Matzen et al., 1994). For instance, both Vodegel Matzen et al. (1994) and Carpenter et al. (1990) could demonstrate that the number of rules explains around 50 percent of item difficulty.

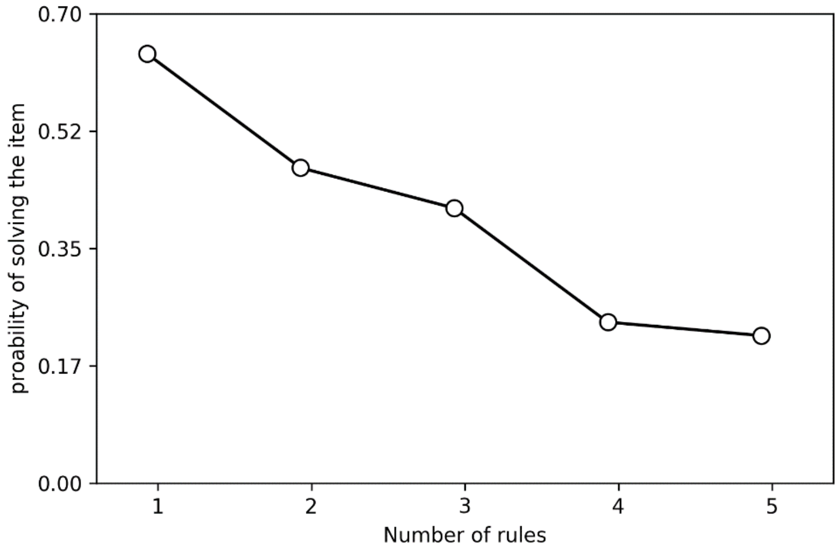



*Figure 6*. Illustration of item difficulties depending on the number of applied rules in a matrix. Plot based on the data by Becker et al. (2016).

To demonstrate that the difficulty in an item depends on the number of rules, the item difficulties of items with one to five applied rules are displayed in *Figure 6*. Although this is only a descriptive visualization of the item difficulties, it illustrates that the probability to solve an item monotonically decreases while the number of rules increases. What is particularly striking is the drop in probability from one rule to two rules, which indicates that there is a qualitative difference between one rule and multiple rules. This raises the question: *why* do multiple rules lead to a higher difficulty?

# 4 Processing in Matrix Reasoning

In order to answer this question, it makes sense to take a closer look at the solution process of matrix reasoning in order to describe on which process these characteristics can have an influence. The goal is to describe which processes can be distinguished when one rule has to be solved and which processes are demanded in addition when multiple rules are solved.

We will first explain three models of geometric analogies since process models on matrix reasoning were derived from them. Following these explanations, we will outline the well-established model for matrix reasoning by Carpenter et al. (1990) and describe the processes involved in solving matrix reasoning items. Finally, we will discuss which of these processes is differentially demanded when applying one or multiple rules in a matrix.

## 4.1 Models on Geometric Analogies

An example item of a geometric analogy is displayed in *Figure 7*. The solver has to answer the question: "`A is to B, as C is to` **?**". The solution process to solve a geometric analogy can be described with three main steps, which are based on the models of Evans (1968) and Sternberg (1977) and was extended by Mulholland et al. (1980).

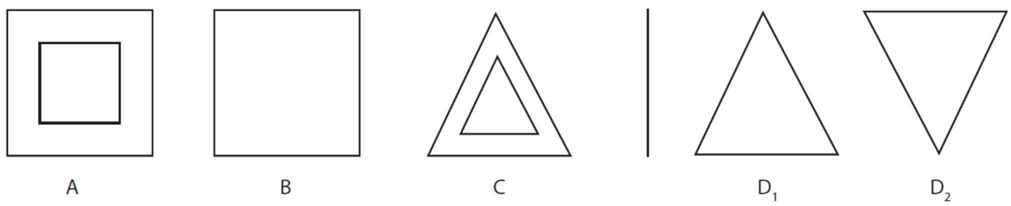



*Figure 7*: Example of a geometric analogy (adapted from Sternberg, 1977)

The first step is an *encoding* process, in which the features of each geometric figures have to be decomposed into single features and encoded. For figure A, one would recognize a small square embedded in a larger square. These features are translated into an internal representation, which is stored in working memory.

The second step is a *comparison/induction* process, in which the solver compares differences between the features. First, a rule X is formulated that accounts for the changes between A and B. In this example, a small square is present in figure A but is absent in figure B. Next, a rule Y, which describes the differences and similarities between A and C. In *Figure 7*, a smaller square is embedded in a larger square in A, and a little triangle is embedded in a larger triangle in C. Thus, a rule can be formulated that describes that "a geometric shape is embedded in a larger geometric shape of the same kind".

The third step is the *application* of rule X. In this step, the rules X and Y were applied to generate a suitable answer for D', which is an image of the (potential) correct answer. Subsequently, D' will be compared to the given response alternatives, and the answer (D) will be selected that corresponds with the mental representation of D'. If there is no D that is fitting to D', then the most fitting D will be selected.

## 4.2 Matrix-Reasoning Processing

The models on geometric analogies can be extended for solving matrix-reasoning tasks. Probably the most established theory on matrix reasoning was developed by Carpenter et al. (1990, see *Figure 8*). This model strongly oriented towards the processes encoding, comparison, and application posited in geometric analogies.

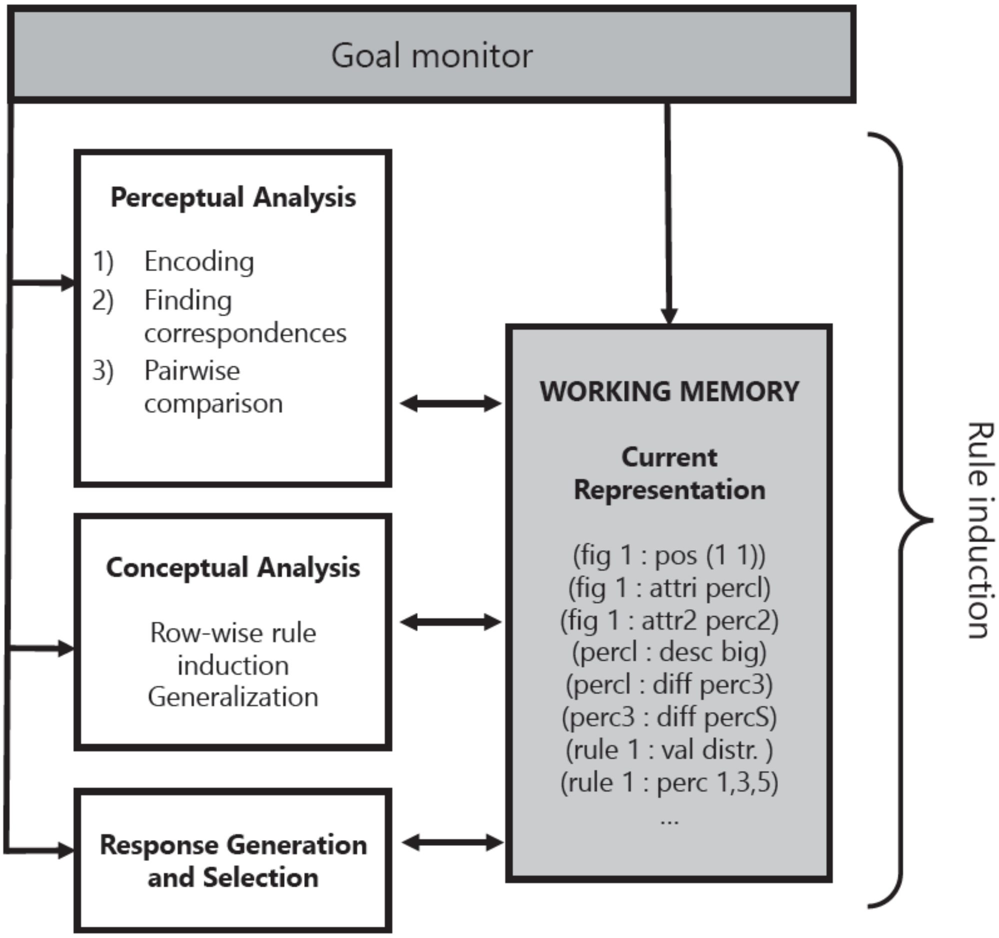



*Figure 8*. Process model of matrix reasoning adapted from Carpenter et al. (1990).

The model by Carpenter et al. (1990) was primarily influenced by computer simulations, which were underpinned by empirical results of verbal protocols, eye movement data, and error patterns. They simulated the solving process of two types of solvers in their model: normal and good performers. The models had structural differences and the evaluation of the solving success of each model was used to derive an understanding of human processing in matrix reasoning. The model for normal performer was named FAIRAVEN, and the model for good performer was named BETTERAVEN. Critical to their model is that they make a qualitative difference in solving one rule compared to multiple rules.

## 4.3 Single-Rule Processing

According to the model by Carpenter et al. (1990), a rule induction process is required to solve *one rule*, which can be divided into certain interim stages: perceptual analysis, conceptual analysis, and response generation (see *Figure 8*). During *perceptual analysis*, the displayed mental representations of the visual material are created and possible groupings of elements are detected, which is described by Carpenter et al. (1990) as

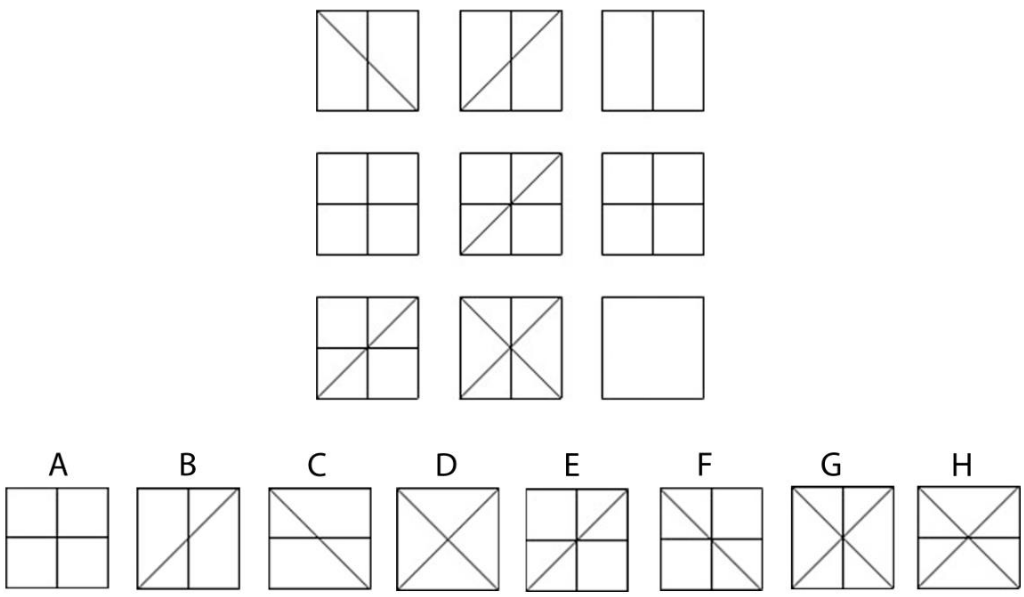



*Figure 9*: Illustrate of an item with one rule (adapted from Feldbrügge, 2012).

*correspondence finding*. In the example item in *Figure 9*, the lines can be summarized to one perceptual group.

The corresponding elements are stored as prepositional lists of elements and values in working memory. For instance "`vertical line in cell 1, vertical diagonal line in cell 1,`…" (see *Figure 9*). The *working memory module* is necessary to store all intermediate representations and results, which was already outlined by Mulholland et al. (1980). Hence, the working memory module acts like a mental sketchpad that stores all important information in a suitable format. Notably, Carpenter et al. (1990) stated that the perceptual analysis is only serving as input for the rule induction and does not necessarily represents a source of interindividual differences.

During *conceptual analysis*, the respondent uses the representation list from the first stage to that each element group follows and *abstracts* this to a conceptual level, i.e. the respondent infers an underlying rule, which is called *rule induction*. Notably, for inducing the rules, the respondent has to find similarities and differences amongst elements in the different cells by *pairwise comparisons*.

In *Figure* 9, only lines presented in the first *and* the second cell, are presented in the third cell. Thus, these elements are governed by an *intersection* rule. During the conceptual analysis, the respondent induces the rules systematically starting in the first row and trying to validate and abstract the rule in the second row. Finally, the rule can be applied to the remaining row containing the solution field, and the answer can be selected from different response alternatives in the response bank (*response generation*). In the example in *Figure* 9, the correct answer would be C. Although response generation is an essential process in other models (e.g., Bethell-Fox, Lohman, & Snow, 1984; Vigneau et al., 2006), we did not address this process in the current work.

Highlighting pairwise comparisons for rule induction is strongly related to the described comparison processes already introduced by Evans (1968), Sternberg (1977) or Mulholland et al. (1980). It emphasizes that an active exploration of the problem is essential to gain an insight into the problem and finally to induce the rule. When generalizing this idea, it can be concluded that the perception of differences and similarities can be seen as one of the core mechanisms when solving reasoning tasks, such as geometric analogies or matrix reasoning tasks, which is reminiscent of the "eduction of relations and correlates" by Spearman (1927) and therefore, underlines the crucial role of pairwise comparisons as an essential mechanism in matrix reasoning.

## 4.4 Multiple-Rule Processing

According to Carpenter et al. (1990), rule induction is also required in order to solve an item with multiple rules. However, solving items with multiple rules differs in several aspects from solving an item with one rule. Carpenter et al. (1990) primarily emphasized a *goal management* process that allows for efficient coordination of multiple rules. However, there are also indications from different studies and models that other processes, such as storing partial solutions and selective encoding, also play a crucial role when solving multiple rules. Interestingly, several studies assume a crucial role of WMC for these processes in items with multiple rules but direct

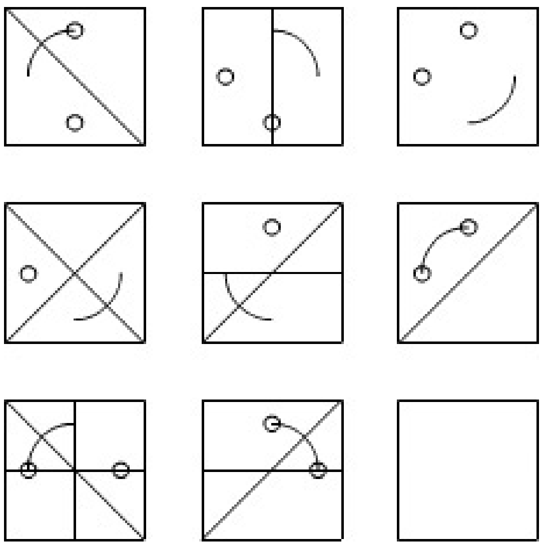


*Figure 10.* Illustration of an item with multiple rules. For simplification, response alternatives are not displayed. Adapted from Feldbrügge (2012).

evidence demonstrating the involvement of WMC is scares as well as a specific definition of WMC. We will briefly review the processes involved in multiple-rule items along with possible implications for an influence of WMC. Please note that we will only roughly introduce the influence of WMC on these processes as indicated in the matrix-reasoning literature. We will more elaborate about the differential involvements of specific aspects of WMC on these processes in the subsequent chapter.

#### 4.4.1 Goal Management

The most established process that is discussed in items with multiple rules is *goal management*, which is a super-ordinate control system that monitors or supervises the solving process. The need for a control system that monitors the solutions process was already highlighted in geometric analogies (Sternberg, 1977).

Carpenter et al. (1990) captured this idea and implemented a goal monitor in their computer model, which was responsible for goal management. In order to successfully solving an item with multiple rules, the item first has to be segmented into sub-goals. Subsequently, the goal monitor ensures that the goals are processed serially to prevent the solver from intermixing the rules, which is also known as *keeping track*. For instance, in *Figure 10*, three rules are applied: one on the lines, one on the small circles and one on the circle segments. Goal management ensures that priority is given to the processing of these rules and that these rules are induced one after the other. Here, it is ensured that all rules are *processed serially* and no rule is forgotten or rule principles are intermixed.

In their simulation studies, Carpenter et al. (1990) found that more complex items (i.e. more rules) are solved when a "goal monitor" is implemented compared to a model, in which this module is absent. In addition, the authors could demonstrate that the performance in matrix-reasoning was strongly related to the performance in the Tower of Hanoi – a task that requires the building and managing of goals. They interpreted this strong correlation as evidence that managing sub-goals is also required in matrix-reasoning items.

Due to the importance of goal management in the model by Carpenter et al. (1990) and due to the substantial correlation of matrix reasoning with WMC (e.g. Unsworth & Engle, 2005), goal management was considered to be *the* process associated with WMC in various studies (Embretson, 1995; Loesche et al., 2015; Unsworth & Engle, 2005). The rationale was, when goal management demands in items with multiple rules exceeds the individuals WMC, the item cannot be solved correctly as the solver loses the track of the solving process and can no longer supervise the goals efficiently. However, the description how WMC should facilitate goal management is quite vague in the literature. Based on the considerations about goal manamgent, we suggest that WMC is resonsible for the storage of sub-goals and a redirection to these stored information while the induction of other rules.

### 4.4.2 Partial solutions

As already described, the working memory acts like a mental sketch-pad, which stores all relevant information for further processing. This also includes the storage of partial solutions of the problem. For example, in the item from *Figure 10* the partial solution of the lines must be stored when the rule of the circle segments is induced.

In fact, there is evidence that the storage of partial solutions is especially crucial for higher-order cognition, such as mental arithmetic (Hitch, 1978) or reading comprehension (Just & Carpenter, 1992). For instance, Just and Carpenter (1992) suggested that “when the task demands are high enough to strain capacity, individuals with a smaller WMC should be less able to perform computations quickly or store intermediate products (p. 143).”

Moreover, the direct association of WMC with partial solutions was also addressed in studies about geometric analogies with multiple rules. Mulholland et al. (1980) described that the number of transformations to a single element place a heavy burden on working memory. They described, that each rule (termed element-transformation) "requires individual placekeeper or slot in working memory" (Mulholland et al., 1980, p. 282). Hence, this implies that the more rules are involved, the more partial solutions have to be stored, which requires more WMC.

Interestingly, although the influence of storing partial solutions is quite often assumed in literature there is less *direct* evidence whether this is actually the case. In addition, the influence of WMC on this process is only described in theory. The current work was set out to provide evidence whether storing partial solutions is an essential process in matrix reasoning or not and whether this is related to WMC.

#### 4.4.3 Selective encoding

The encoding process is another process that is also demanded in items with multiple rules. It could be demonstrated that items with multiple rules are visually more complex as they consist of several (overlapping) elements. To solve an item with multiple rules, the figure has to be segmented and only those elements that are relevant for the current rule must then be encoded (Meo et al., 2007; Primi, 2002). For instance, only the lines have to be encoded when inducing the underlying rule of the lines while other elements (small circles and circle elements) have to be ignored (see *Figure 10*).

Primi (2002) argued that the demands of selective encoding and goal management are traditionally confounded in matrix-reasoning items. That is, items with multiple rules always require selective encoding of the current rule *and* goal management of the applied rules. To disentangle both selective encoding and goal management demands, Primi (2002) constructed two versions of matrix-reasoning tests in which these two demands were independently manipulated.

In the first version, all elements and features were relevant. In contrast, in items of the second version, irrelevant attributes were added to relevant elements or the elements were re-arranged in every row. For the example in *Figure 10*, this could mean that irrelevant colors or shades are added to the lines or the circles are re-arranged to different positions in each cell.

Primi (2002) found that adding irrelevant attributes is the main determinant of item difficulty besides the number of rules in an item. He argued that in items with irrelevant attributes the respondent is distracted by irrelevant groupings and selective encoding of relevant attributes is mandatory. Primi (2002) suggested these irrelevant groupings disrupted the perceptual continuity and as a consequence, the creation of a stable mental representation of the relevant elements was hampered. Consequently, an underlying rule could not be induced properly.

Primi (2002) argued that a selective encoding mechanism that is related to working memory is essential to ensure that only relevant elements are encoded and irrelevant elements are ignored. This result has two important implications: First, it emphasizes the importance of successful selective encoding besides goal management. Since in both item versions the number of rules is constant, goal management requirements are also constant and the effect of the higher difficulties in items with irrelevant attributes can be only attributed to the requirements of selective encoding. Second, it suggests that a not more specified aspect of WMC is needed to selectively encode relevant elements and ignore irrelevant elements.

More importantly, the demands of selective encoding cannot only explain why items with *artificial* added irrelevant attributes are harder to solve. It also indicates why items with multiple rules are harder to solve as these items also meet the requirements selective encoding. According to the model of Carpenter et al. (1990) rules are processed serially. Hence, when processing the first rule, elements from other rules have to be ignored as these elements are irrelevant for the current rule. In line with Primi (2002), the irrelevant elements of the rules, which are currently not being solved, could also disrupt the perceptual continuity of the currently processed rule.

However, the description of perceptual continuity is quite vague and it remains unclear *how* the irrelevant elements hamper processing during the rule induction. In this work, we assumed that the segmentation of relevant and irrelevant elements for processing the current rule is time-consuming. This can be supported by a recent study, demonstrating higher response times for item with multiple rules compared to items with one rule (Becker, Schmitz, Göritz et al., 2016). Longer response times could indicate that the pairwise comparisons of cells in the matrix are disturbed, so that the rule is slower or maybe also not correctly solved. Hence, we consider the disturbed perceptual continuity in items with multiple rules as *disturbed flow* of the pairwise comparisons.

This assumption that multiple elements in a matrix hamper the smooth processing of the rules can be supported by the study by Meo et al. (2007). The authors found evidence that matrices items were more difficult when the elements were harder to identify and were overlapping. Meo et al. (2007) concluded that these items prevent the isolation from other elements and therefore, the creation of economical appropriate representations in working memory. The authors suggest that "these findings alert us to the possibility that perceptual factors do have a substantial part to play in the solution of Raven's Progressive Matrices, as well as working memory capacity, and that individual differences in people's ability to identify item elements may be an important source of variance in test scores." (Meo et al., 2007, pp. 367–368).

In summary, there are reasonable considerations that one potential source of difficulty lays in selective encoding of relevant elements since items with multiple rules usually contain overlapping groups of elements. When processing the current rule, elements of other rules have to be ignored as these elements are irrelevant for the current rule and disrupt the visual processing. However, based on the previous studies it remains unclear to what extent selective encoding is involved besides goal management since both demands are traditionally confounded as items with multiple rules theoretically always require selective encoding and goal management. In addition, it is an ongoing question how irrelevant elements hamper the rule induction but a disrupted perceptual continuity during pairwise comparisons seems to be a promising candidate to answer this question.

# 5 Interim Conclusion 1

Taken together, matrix reasoning is a paramount tool to assess gF as it requires the detection of underlying rules in novel problem, which has been posited to be core abilities in gF (Spearman, 1927; Thurstone & Thurstone, 1941; Horn & Cattell, 1966). Interestingly, the number of rules is the main predictor of item difficult, which implies that a source of intelligence lies in the ability to deal with a large amount of information in matrix-reasoning tasks. The reasons for this can be found in different processes during the solution of a matrix-reasoning item. The first process is goal management, which is an efficient handling of the rules and supervision of the solving process. The second process is the storage of partial solutions. The more rules are applied in a matrix, the more intermediate information of the solution has to be stored. The third is selective encoding. Since the rules are processed serially, elements, which are irrelevant for the current processed rules have to be ignored and only the relevant elements have to be encoded selectively.

Notably, all of these processes are somehow associated with WMC in the literature. First, goal management requires processes to control attention towards goal-relevant information and the redirection to already stored material. Second, storing partial solutions is associated with the storage capacity of working memory as every partial solution requires an "individual placekeeper" (Mulholland et al., 1980). Third, WMC is needed to selectively encode only relevant elements in items with irrelevant elements or multiple rules.

However, none of the studies about matrix-reasoning processing have directly assessed WMC, so that the connections between the processes and WMC are only based on theoretical considerations. Moreover, process models on matrix reasoning are mainly influenced by computer simulations. To our knowledge, there exists no *empirical* study that has considered the processes of matrix-reasoning individually and observed the role of WMC on these processes. Hence, the aim of the present work is to isolate the processes *goal management*, *storage of partial solutions* and *selective encoding* by experimental manipulations and observe the influence of WMC on these processes. As the described processes imply a qualitatively different involvement of WMC, we will focus in the next chapter on the concept and measurement of WMC to provide a deeper understanding of different aspects of WMC considered in the present work.

# C Working Memory Capacity

In general, *working memory* refers to a cognitive and neuronal system that enables an active maintenance of information for further processing (Zimmer, 2008). Working memory *capacity* (WMC), on the other hand, is defined from different perspectives, which influences the understanding of the contribution from WMC to gF. Consequently, these perspectives must be taken into account in order to describe the impact of WMC on the described matrix-reasoning processes on a more functional level.

Since it is important for this work *why* WMC facilitates the described matrix-reasoning processes, we would like to focus on certain aspects of WMC that seem promising to be related to these processes. Hence, in the following chapter we will outline a theoretical model of working memory and how WMC can be defined in this model. At this point, we will focus on two different perspectives: one considers WMC as storage capacity, the other as controlled attention. Subsequently, we will describe how these aspects of WMC can be measured and how they are related to gF, in order to derive why these aspects are meaningful for the matrix-reasoning processes described above.

## 1 The Concept of Working Memory

One of the most-established models about working memory is the tripartite model by Baddeley and Hitch (1974). In this model, the working memory system consists of three main parts: two slave-systems – the phonological loop and the visual-spatial sketchpad – and the central executive. Each slave-system represents a capacity-limited temporary memory for a given material and both systems are independent of each other. The *phonological loop* can maintain verbal information for a few seconds without rehearsal, and the *visual-spatial sketchpad*, on the other hand, provides a temporary storage for spatial and visual information. The *central executive* is a capacity-limited control system that monitors the processes of the phonological loop and the visual-spatial sketchpad. This component ensures that the resources are deployed to peruse the current goals, which can be achieved for instance by directing or dividing attention when it is required.

The view of working memory as a system that integrates both storage and control components was mainly influencing further models of working memory. One prominent model integrating both storage and control components is the *embedded process model* by Cowan (1988; 1995; see *Figure 11*).

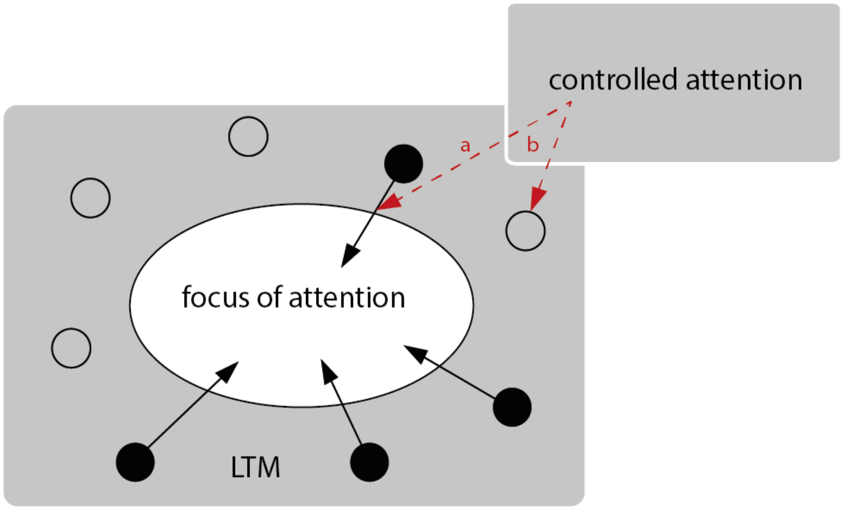



*Figure 11.* Illustration of the embedded process model by Cowan (e.g. 1988)

The embedded process model is less modular than the tripartite model by Baddeley and Hitch (1974) but focusses more on the *processes* of working memory. That is, Cowan refrains from specifying different components that are specialized for information from different modalities (i.e. verbal vs. visuospatial). On the contrary, there exist no separate memory systems for different modalities but all information is governed by the same processes. At this, Cowan (1988, 1995) describes the memory system in two stages: the first stage is long-term memory (LTM) and the second stage is the *focus of attention* that is embedded in an activated part of LTM.

All information, which is voluntarily retrieved from LTM or encoded from the surrounding are entering the focus of attention. The focus of attention can, therefore, be considered as a "spotlight" that adjusts its light on the information that is considered for cognition.

To bring the information within the focus of attention – or with respect to the metaphor: to adjust the spotlight – a voluntary process is necessary. This voluntary process is associated with "*control of attention*", which can be described as the direction of attention to the stimuli that are intended to be processed or as the blocking of irrelevant information. In *Figure 11*, the direction of attention to relevant stimuli (path a) and the blocking of irrelevant features that are not supposed to enter the focus of attention are displayed as two examples of the functioning of controlled attention (path b).

## 2 What is Working Memory Capacity?

Based on the model of Cowan (e.g. 1995), two different definitions about WMC have emerged, and these definitions were mainly influenced by two different views on the model by Cowan (1988; 1995). One view considers the size of the focus of attention as WMC and therefore, WMC is defined how much information can be actively maintained under voluntary control (Cowan, 2001; Cowan et al., 2005; Luck & Vogel, 1997). The amount of maintained information is often termed as the *scope of attention* (e.g., Cowan, 2001) or as *storage capacity* (e.g., Chuderski, Taraday, Nęcka, & Smoleń, 2012), and in this work, we will refer to the term storage capacity.

Another view describes WMC as the *effectiveness* of controlled attention that brings or keeps the relevant information into the focus of attention (Engle et al., 1999; Engle, 2002). Otherwise, information would interfere with other information and therefore, could not be used for further cognition. Hence, this WMC definition is not about the number of maintained items in the focus of attention but about the control processes that enable a successful activation of information within the focus of attention. Hence, the amount of information within the focus of attention can be considered as a result or a positive side effect of controlled attention.

In the next sections, we will describe the main characteristics of both storage capacity and controlled attention and describe how they are commonly assessed. We will take up this idea of several definitions or aspects of WMC for describing possible influences on the described matrix-reasoning processes. However, before deriving implication of these definitions for matrix reasoning, we will describe how these aspects are traditionally measured and how they are related to gF.

## 2.1 Storage Capacity

To estimate how much independent information can be possibly held active within the focus of attention, tasks have to be utilized, which minimize strategical techniques such as chunking or rehearsal (Cowan, 2001). One prominent task for estimating storage capacity is the change detection paradigm (Luck & Vogel, 1997; Phillips, 1974), which is sometimes also termed as visual arrays (e.g., Cowan et al., 2005).

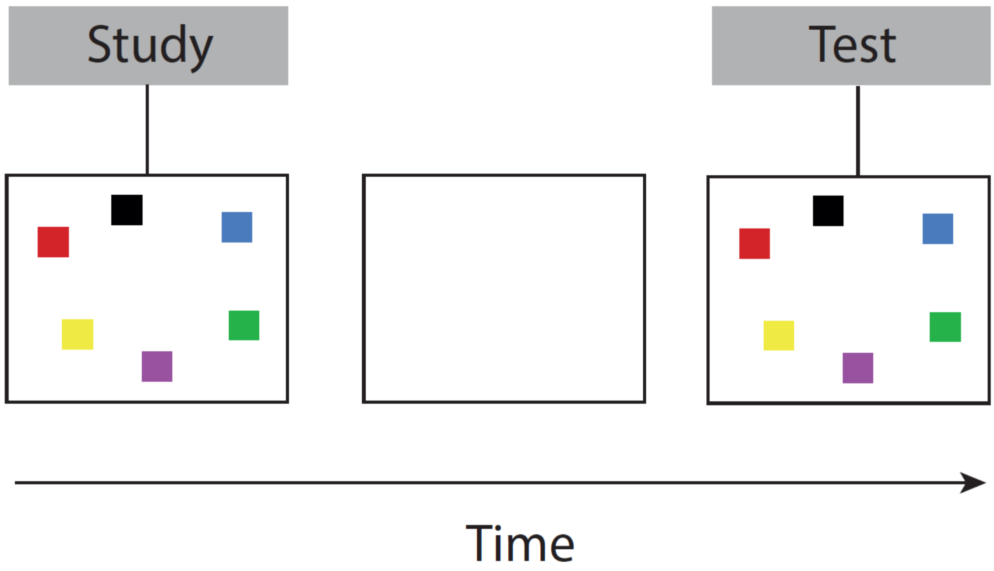



*Figure 12*: Illustration of a change detection task with color change. Displayed is a no-change trial.

In a change detection task, (see *Figure 12*) a sample of items is briefly shown for a few hundred milliseconds in the study phase and re-presented after a short delay in the test phase. The sample in the test phase is identical or differs in some aspect (e.g., color or shape of one or more items is changed). Individuals are required to judge whether the display in the test phase is the same as in the study phase or not. By varying the number of displayed items, individuals' number of simultaneously retained items can be estimated. This number is referred as K-index and considered as an estimate of storage capacity, which can be calculated with the formula K = set size × (hit rate − false alarm rate) (Pashler, 1988, Cowan, 2001). K represents the number of stored items in memory, *set size* the number of presented items in the sample display, *hit rate* the proportion of correctly detected changes, and *false alarm rate* the proportion of given change responses to non-change trials.

Several studies reported a mean storage capacity, estimated by the change detection task, of about three to four items (Cowan, 2001; Luck & Vogel, 1997; Vogel & Machizawa, 2004). More interestingly from a differential perspective, the individual storage capacity differs between individuals ranging from 1.5 to 6 items (Vogel & Awh, 2008). This could also be shown in electrophysiological studies, which demonstrates that there is a neural correlate of inter-individual differences in storage capacity (Vogel & Machizawa, 2004).

## 2.2 Controlled Attention

Although the estimation of storage capacity seems straightforward, this is not the case for controlled attention. First, two different types of tests of controlled attention have to be distinguished. One type of tasks assesses several processes of controlled attention *without* a working memory task, and another type assesses controlled attention in a working memory task. Therefore, we term these two types *memory-unrelated* tasks and *memory-related* tasks for the remaining chapters. Although we were not focusing on memory-*unrelated* tasks in the current work, we will shortly review some of these tasks as the main characteristics of these tasks have important implications for the processes of controlled attention in memory-*related* tasks.

### 2.2.1 Memory-unrelated controlled attention

Memory-unrelated tasks require the direction of attention to a certain task goal, especially in face of distractors or interference. Hence, these task test how efficiently one can protect a task goal from a challenging dominant process. One example is the Stroop task (Stroop, 1935), in which color words are presented (e.g. red or green). Importantly, in congruent trials, the words were sometimes presented in the same color as the wording (the word "red" is presented in red). In incongruent trials, on the other hand, the words are presented in a different color as the color word (the word "red" is presented in green). The task of the participant is to report the colors in which the words are presented. When the color of the word and the word are incongruent the participants respond less accurate or with longer response times compared to congruent conditions. The reason is that the reading of a word is a habituated or an automatic response that stands in conflict with the current task goal (name the color in which the word is written in). The reading of the words interferes with the actual goal, which causes a longer response time as the participants have to disengage from the reading. Alternatively, if the disengagement fails, the participants read the word and give a false response.

Another task for memory-unrelated tasks is the flanker task (e.g. Heitz & Engle, 2007). In one variant of this task, five arrows are presented in a row and the participant is asked to indicate whether the middle arrow is pointing to the left or to the right direction. In congruent trials, all arrows are pointing to the same direction as the middle arrow, and in the incongruent trials, the surrounding arrows pointing to the opposite direction than the middle arrow. Hence, in incongruent trials, participants have to ignore the distracting arrows and control the focus on the middle arrow to ensure that the answer about the direction of the middle arrow is given properly.

In all, these tasks demonstrate that controlled attention is required to direct attention to task-relevant goals, which taps into different processes such as blocking of interference, combating interference or redirecting to current task goals.

As memory-*unrelated* tasks of controlled attention, memory-*related* tasks also require to focus on a current goal in the face of interference or distraction. However, in contrast to the memory-unrelated tasks, the goal is not to perform an action (e.g. indicate the direction of an arrow) but to maintain information in working memory. In terms of the model by Cowan (1988, 1995), the goal is to hold information active in the focus of attention in the face of distracting or interfering information, and controlled attention is responsible that this goal can be successfully achieved. One construct, which covers several of the processes of controlled attention is *storage and processing*, which requires the storage of information in working memory while processing a competing secondary task.

### 2.2.2 Storage and processing

The most common task set to assess storage and processing are complex span tasks. A complex span task consist of two tasks, which are alternately presented. The first task is a storage task, and the second task is a processing task. One prominent example of complex span tasks is the Operation Span (Ospan; Unsworth, Heitz, Schrock, & Engle, 2005), in which letters from a list have to be stored (storage task), and between each presentation of a letter, a math operation (processing task) has to be solved (*Figure 13)*. The number of switches between the letter presentation and math operations depends on the current list length, which is usually between two and seven. In *Figure 13*, a list of the length two is applied, since two letter presentations and two math operations are alternately presented.

In a subsequent recall phase, the participants are required to indicate the serial order of the presented letters. For instance, in *Figure 13*, first the F and then the P has to be indicated. The number of correct recalled letters in their serial order is taken as an indicator for WMC, which is known as the partial scoring method (for a review of the scoring methods see Conway et al., 2005).

It is important to note that the performance of the processing task is not directly taken into account when calculating WMC in complex span task. However, only items are considered for the score, in which the performance level was above a certain threshold, which is usually around 85 percent (see Unsworth et al., 2005). This controls that participants actively performed the processing task, and not only retained the letters of the storage task. For this example, this would mean, that 85 percent of the math operations have to be performed successfully to ensure that this item can be included for calculating WMC.

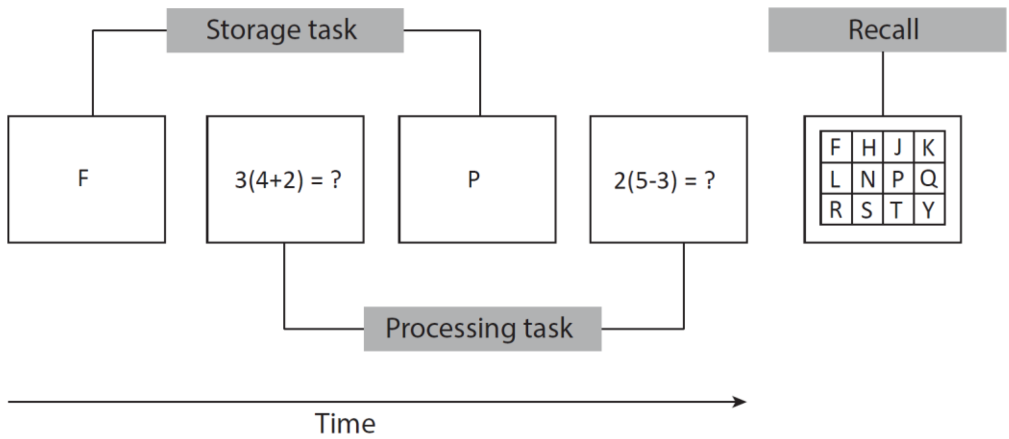



*Figure 13*: Procedure of the Operation Span (Ospan) as an example of complex span tasks for assessing *storage and processing*. Illustration modified from Foster et al. (2014).

Several studies provided evidence that complex span tasks tap into a broad range of processes assessed in memory-unrelated controlled attention. Substantial correlation between complex span task and anti-saccadic task (Unsworth, Schrock, & Engle, 2004), Stroop task (Kane & Engle, 2003), and flanker task (Heitz & Engle, 2007) indicate that the control of attention towards relevant goals is common in both memory-unrelated and memory-related tasks of attentional control. In detail, these studies demonstrate that complex span tasks, covering several processes such as the direction of attention, blocking of interference, and combating interference, and thus qualifies them as a paramount tool to assess controlled attention in a working memory context (Conway et al., 2005).

*Figure 14* illustrates that storage plus processing, assessed by complex span tasks covers most of the controlled attention processes (see also Unsworth et al., 2014). A specific process of controlled attention is the blocking of irrelevant information, which receives a great deal of attention in research on WMC, but is not specifically captured by storage plus processing (Cowan & Morey, 2006; Vogel, McCollough, & Machizawa, 2005). Since selective encoding and blocking of irrelevant matrix elements have been identified as an important process in matrix-reasoning, we want to present the filter task in the next section that measures the ability to avoid irrelevant stimuli entering working memory.

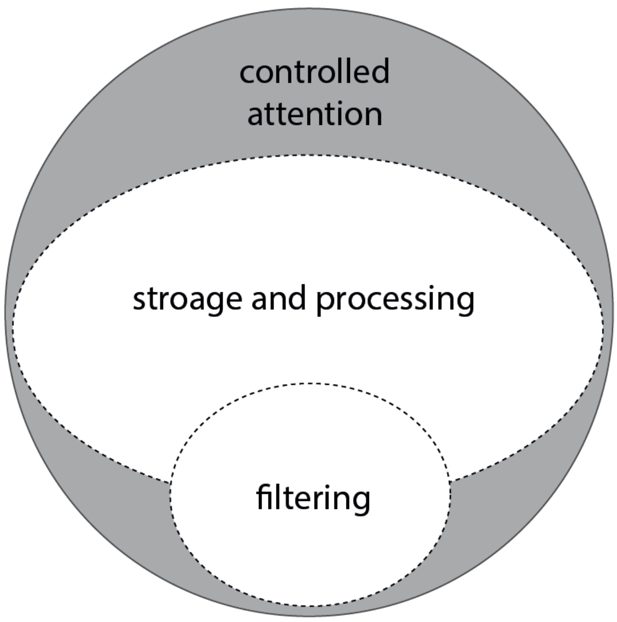



*Figure 14*. Storage and processing and filtering as several aspects of controlled attention.

### 2.2.3 Filter task

The filter task is a modified version of the change detection task in which information has to be maintained in the face of irrelevant, distracting information (e.g. Vogel et al., 2005). The procedure of this task is nearly identical to the standard change detection with the only difference that on some trials distractor-items are presented in the memory array in addition to the to-be-remembered, relevant items. An example task is presented in *Figure 15*, in which the color of the squares has to be maintained but colored rectangles are presented in addition as distractor-items. Importantly, the participants are asked to remember only the critical feature of the relevant items (i.e., color of the sqaures), and only a potential change in the relevant items has to be indicated in the test array.

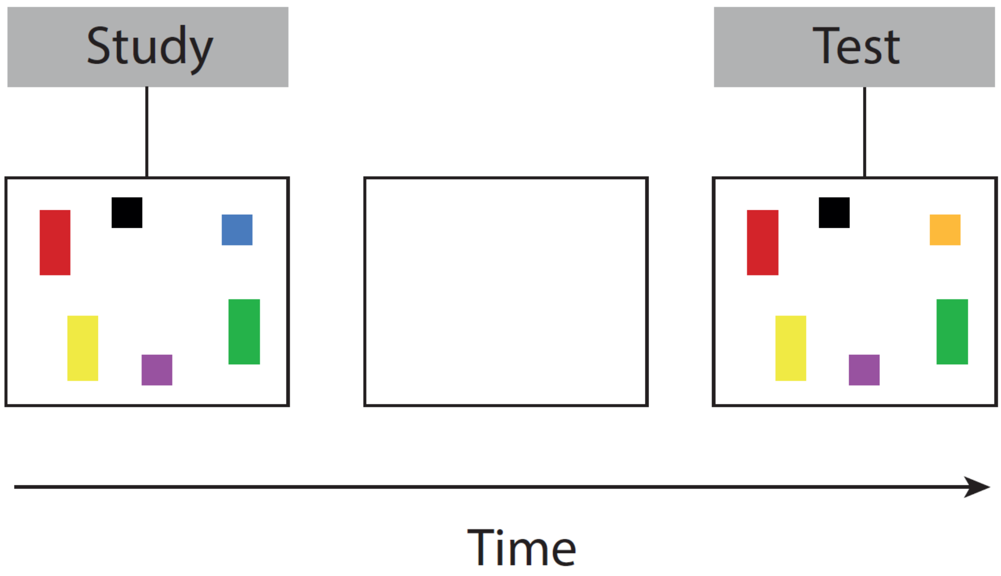



*Figure 15*: Procedure of the filter task (adapted from Liesefeld et al., 2014). Displayed is a change-trial.

Since the presentation of all items in this example would exceed the storage capacity of the participants (cf., Cowan, 2001), they cannot simply retain the color of all items (squares and rectangles) but have to ignore the color of the rectangles and *filter out* the squares. In other words, they have to avoid that the rectangles are entering the in the focus of attention and therefore, contaminate working memory with distractor-items. To estimate the efficiency of this filtering process, distractor-present (relevant + distractor items) trials are presented in addition to distractor-absent trials (only relevant items) and the performance in these two conditions are compared. When the performance decreases from distractor-absent to distractor-present condition, this can be seen as an indicator that the filtering process was not as successful as irrelevant information were capturing space within the focus of attention and therefore, less capacity was left for the relevant information. On the other hand, if performance is similar in both conditions, filtering was successful. The difference between distractor-absent and distractor-present can, therefore, be interpreted as filtering costs since this difference describes how much the performance decreases when distractors are presented.

An overall decrease of performance in distractor-present trials compared to distractor-absent trials is well-reported (Lee et al., 2010; Spronk, Vogel, & Jonkman, 2012). Importantly, it could be demonstrated that filtering efficiency differs between individuals (e.g., Liesefeld, Liesefeld, & Zimmer, 2014; Vogel et al., 2005).

In order to enrich the understanding of what distinguishes people with good and bad filter efficiency, two aspects can be considered in detail. The first is the relatedness of the filtering task to the flanker task. Machizawa and Driver (2011) compared filter efficiency with performance in the executive control task of the Attention Network Test (ANT), which is one variant of the flanker task. Using a principal component analysis, they found evidence that filtering efficiency and executive control of the ANT are loading on the same factor. This suggests that both the filter task and the flanker task are related to a similar attentional control processes to ignore irrelevant information in order to protect a certain goal. In terms of filtering efficiency, this means that individuals are more efficient in preventing irrelevant information that is entering the focus of attention.

The second aspect is attentional capture. The question is raised: do individuals with better filtering capabilities more efficiently *ignore* irrelevant information or are they more capable of *disengaging* from irrelevant distractors? Evidence supports the second hypothesis showing that the initial attention of all individuals is captured by irrelevant information but that more able respondents can disengage earlier from irrelevant information and redirect attention to relevant information (Fukuda & Vogel, 2009, 2011). In sum, these studies provide evidence that filtering is associated with an efficient mechanism for fast attentional capture from irrelevant information.

## 2.3 WMC in a Nutshell

In conclusion, WMC comprises two different aspects: storage capacity and controlled attention (see *Figure 16*). Storage capacity describes the number of independently maintained items within the focus of attention and is assessed by the change detection paradigm. Controlled attention, on the other hand, describes several processes that ensure an efficient encoding, retrieval, and maintenance of relevant information. In terms of memory-related tasks, one aspect that covers several of these processes is storage and processing, which is commonly assessed by complex span task. One specific process is filtering, which describes the process of selectively

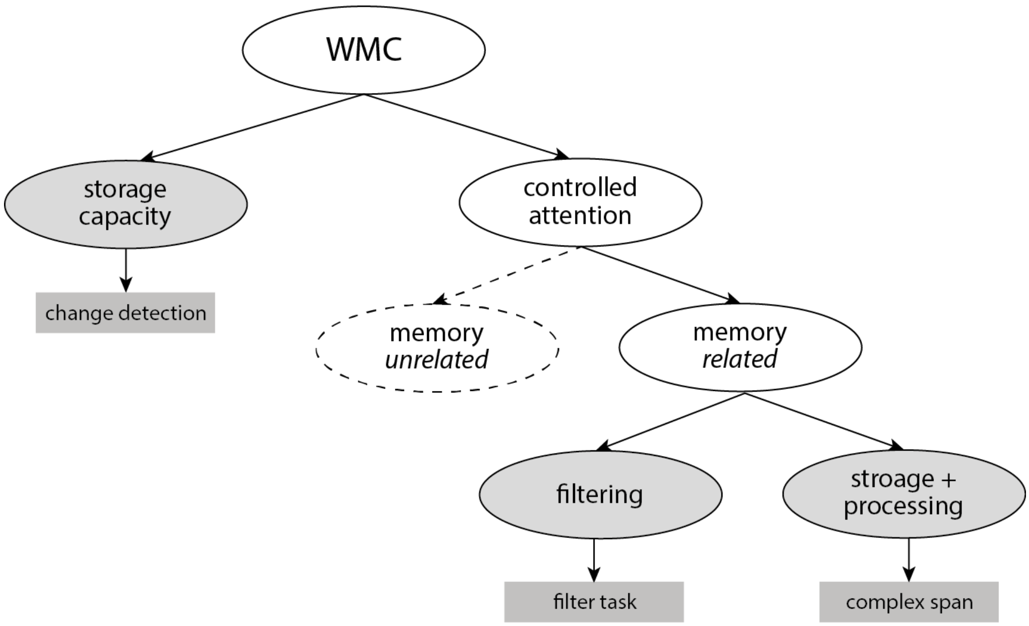



*Figure 16*. Different perspectives on WMC and its measurement.

encoding relevant information while preventing irrelevant information entering the focus of attention.

In the next chapter, we will outline how these different aspects of WMC are related to gF in order to gain an insight as to how they could be related to the matrix-reasoning processes relevant for the current work.

# 3 WMC and gF

Before we discuss the relationship between WMC and gF, we will briefly describe how this relationship can be quantified and described. Studies that investigated the relationship between WMC and gF usually applied latent-variable approaches, which describe correlational relationships between WMC and gF via structural equation modeling (SEM; e.g., Hilbert & Stadler, 2017). SEM is a combination of confirmatory factor analyses (CFA) and path analyses. In a CFA, the shared variances between different tasks of the same construct can be reduced to one common underlying latent variable. As only shared variance between the tasks are extracted and considered in the latent variable, task specificity and measurement error of these tasks are not included in the latent variable and considered as “residuals”. For instance, the shared variance between different gF tests can define a latent gF factor, in which task-specific variances, which result from specific test characteristics (e.g. visual material in one task and verbal material in another task), are not included. After defining latent variables, inter-relations between each of the latent variables can be tested, such as moderations, mediations or unique and common contributions of two constructs on another construct.

Particularly, when examining the influence of WMC on gF, several aspects of WMC, such as storage capacity and controlled attention, are often contrasted. Here, the different aspects of WMC are controlled by other aspects and evaluated whether the remaining variance contributes to the prediction of gF.

Based on the considerations of Cowan (e.g. 1995), several studies investigated whether storage capacity or controlled attention drives the relationship between WMC and gF. In other words, researchers attempted to discover whether the scope of the focus of attention or the control over the scope is associated with gF. Several studies have demonstrated that both storage capacity and controlled attention share substantial, unique variance with gF (Chow & Conway, 2015; Cowan, Fristoe, Elliott, Brunner, & Saults, 2006; Shipstead, Lindsey, Marshall, & Engle, 2014; Shipstead, Redick, Hicks, & Engle, 2012; Unsworth et al., 2014).

## 3.1 Storage + Processing and gF

Controlled attention, which is especially assessed by storage and processing tasks, is considered as the main aspect to drive the relationship to gF. For instance, Engle et al. (1999) predicted gF with two different types of tasks: complex span tasks, which tap into both storage and processing, and simple span tasks, which only require the short-term storage of information without a competing processing task. The authors extracted the common variance of both complex span tasks and simple span tasks and demonstrated that complex span task but not simple span task share remaining unique variance with gF. They concluded that storing information in the face of a distracting and interfering processing task is essential for gF. In other words, keeping information within the focus of attention while combating interference and distraction was assumed to be most relevant for gF.

The importance of controlled attention was replicated in subsequent studies (e.g., Conway et al., 2002) and underpinned by substantial correlations between memory-unrelated tasks, such as Stroop, and gF (e.g., Buehner, Krumm, & Pick, 2005; Unsworth & Spillers, 2010). This was taken as evidence that controlled attention is involved in “organizing” the information within the focus of attention while performing a secondary task (Shipstead 2014). In terms of gF, this means that rule principle or goals are maintained during another induction process, and attention is redirected in order to integrate all information into a response for the current problem.

## 3.2 Filtering and gF

In addition to the importance of storage and processing, which covers many processes of controlled attention, filtering as one specific process of controlled attention has also been highlighted as an important process in the prediction of gF. For example, Cowan et al. (2006) demonstrated that performance in a task, which requires the filtering of relevant out of irrelevant information (selective listening-procedure) shared unique variance above storage capacity with gF. In addition, others could show that memory-unrelated tasks of controlled attention, which are related to the filter task such as the flanker task were related to gF (Shipstead et al., 2014; Wongupparaj, Kumari, & Morris, 2015). This demonstrates that preventing irrelevant stimuli entering the focus of attention is an important process for the prediction of gF.

However, none of these studies assessed filtering by the variation of the change detection task. Since this task is commonly used when estimating the filtering efficiency in studies of cognitive psychology, it is surprising that this task was not considered in detail when predicting gF.

In fact, to our knowledge, there is only one study which considered the performance of the filter task from a perspective of inter-individual differences. Shipstead et al. (2014) assessed the filter task besides a large quantity of tasks such as the change detection, complex span task, memory-unrelated controlled attention tasks, and several gF tasks. The found that filtering was associated with memory-unrelated tasks, such as the flanker task, which replicates the already described finding by Machizawa and Driver (2011). In addition, they found that the relationship between filtering and gF was mainly mediated by memory-unrelated controlled attention tasks. However, the influence of filtering on gF was not addressed as the main issue in this article but more as a “byproduct” of the analyses.

In addition, one major shortcoming is that the authors included negative K (i.e. negative storage capacity) in their analyses, which could bias the analyses and thus make it difficult to interpret the results.

However, the study has the advantage that it introduced another method for describing filter efficiency. In addition to the traditional method of calculating filter costs (distractor-absent minus distractor-present), Shipstead et al. (2014) control the variance, which a latent storage-capacity factor shares with in a latent filter factor. Since the filter task requires both storage and filtering, the variance of the task that goes back to storage is controlled, and only the variance that is related to filtering remains.

In summary, it could be shown that filtering of relevant information is related to gF. However, there is little evidence whether filtering, measured by the variation of change detection task is related to gF. For the present work it is important to take a closer look at this influence of specific filtering task, as we assume that filtering in the working memory is related to selective encoding in matrix-reasoning. In addition to the usual calculation of filter costs, we also want to consider the alternative method of Shipstead et al. (2014).

## 3.3 Storage Capacity and gF

Besides controlled attention, storage capacity has also be shown to be an important aspect of WMC in the prediction of gF. Using the change detection paradigm, several studies could demonstrate that storage capacity shares a substantial amount of variance with gF even when controlling for controlled attention (e.g., Chow & Conway, 2015; Fukuda, Vogel, Mayr, & Awh, 2010; Unsworth et al., 2014). In addition, it has been demonstrated when controlling a controlled attention factor for storage capacity, the correlation between controlled attention and gF became unstable (Chuderski et al., 2012; Colom, Flores-Mendoza, Quiroga, & Privado, 2005; Colom, Rebollo, Abad, & Shih, 2006), which also indicates that storage capacity is a substantial aspect of WMC driving the relationship to gF.

Although for controlled attention it seems more plausible to describe a direct role for the involvement in gF, this is more complex for storage capacity. For example, the blocking of irrelevant elements can be transferred to the blocking of irrelevant elements or solutions in gF tasks. For storage capacity, it is difficult to find direct evidence of an involvement in gF. However, it is assumed that storage capacity is involved in the representation of hypotheses of the problem, goals, and *partial solutions* (Unsworth et al., 2014). Hence, individuals with higher storage capacity can potentially maintain more information during the solution process.

## 3.4 Implications on Matrix Reasoning

One shortcoming of the described studies on the involvement of WMC in gF is that they are more on a speculative level as direct evidence for these suggestions is scarce. For instance, to our knowledge, no data were provided to show whether storage capacity is in fact involved in maintaining partial solutions in gF tests, or whether controlled attention facilitates the redirection of attention to task-relevant goals during the induction process.

However, the suggestions raised in the articles are highly relevant for the current work as they entail important implications on how WMC could be involved in the three matrix-reasoning processes we have described herein: storage of partial results, selective encoding, and goal management.

First, these studies suggest that storage capacity is involved in maintaining partial solutions and representations of a given problem (e.g. Unsworth et al., 2014). As matrix reasoning with multiple items requires the storage of partial solutions (e.g. Carpenter et al., 1990), it is therefore plausible that storage capacity is involved in maintaining partial solutions in an item until the whole item is solved. As the estimated storage capacity is also considered as the number of "slots" one can maintain in working memory (for a review see Luck & Vogel, 2013) the association with storing partial solutions seems evident as Mulholland et al. (1980) suggested that each partial solutions requires an individual slot. Hence, we considered storage capacity to be involved in storing partial solutions in matrix-reasoning items with multiple rules.

Second, the maintenance of current goals in face of a competing secondary task and the redirection to them in storage and processing tasks corresponds to the same requirements as for goal management. As goal management is required for keeping track of the goals in a matrix-reasoning item, the maintained goals have to be protected against interfering and distracting information during rule induction (e.g., Carpenter et al., 1990). Hence, solving a matrix-reasoning item can be compared to a storage and processing task, in which information has to be maintained while performing a competing secondary task, and attention has to be re-directed to the stored items and recalled in the correct order. Thus, we considered storage and processing to be involved in goal management in matrix-reasoning items with multiple rules.

Third, filtering in working memory, as one specific process of controlled attention, can be linked to selective encoding in matrix-reasoning. Both filtering and matrix-reasoning with multiple items, require to selectively encode relevant information while blocking irrelevant information, and therefore, a common mechanism is that the respondent is not distracted by irrelevant information. As studies on matrix-reasoning suggest that one source of difficulty lies in the demands of selectively encoding relevant information (Meo et al., 2007; Primi, 2002), filtering is a promising process of working memory, which plays a crucial role. In addition, studies on matrix-reasoning indicate that the segmentation of relevant and irrelevant information and the disengagement from irrelevant information is a time-consuming process (e.g., Becker, Schmitz, Göritz et al., 2016), which could also be demonstrated in studies on filtering in working memory (Fukuda & Vogel, 2009; 2011). Hence, we considered filtering to be involved in selective encoding of relevant elements for the current processed rule in matrix-reasoning items with multiple rules.

Before we will outline our studies for testing these relations of certain aspects of WMC and the respective matrix-reasoning processes, we will describe how WMC is considered to be involved in matrix reasoning in the literature. This is necessary, as the involvement of WMC on matrix-reasoning is not as differentiated as in latent-variable approaches investigating the relationship between WMC and gF.

# D Matrix Reasoning and WMC

Although the outlined latent-variable studies on WMC and gF revealed that multiple aspects of WMC, such as controlled attention and storage capacity, drive the relationship, this view is scarcely considered in studies investigating the impact of WMC in matrix reasoning. This is surprising since matrix reasoning is understood as one of the most essential tasks to assess gF, as outlined above, and therefore one would expect that the knowledge of latent-variable approaches is transmitted to matrix reasoning research. In contrast to latent-variable approaches, WMC is mostly described as controlled attention (e.g., Unsworth & Engle, 2005) or as a "composite score" of several WMC tasks (e.g., Loesche et al., 2015) intermixing *storage* with *storage and processing* demands. Hence, these studies described how controlled attention or whether a vague defined WMC can facilitate processing in matrix reasoning. We will shortly review the key studies investigating the role of WMC in matrix-reasoning items with multiple rules along with shortcomings relevant to the current work.

## 1 Multiple Rules and WMC

To uncover whether WMC is more involved in matrix reasoning with multiple items, a correlational approach was applied in several studies (Salthouse, 1993; Unsworth & Engle, 2005; Wiley, Jarosz, Cushen, & Colflesh, 2011). At this, each matrix-reasoning item was correlated to a WMC estimate, and when items with multiple rules require more WMC, items with multiple rules should reveal higher correlations with WMC in contrast to items with only one rule.

However, all studies reported that the correlation between matrix reasoning items and WMC did not vary as a function of applied rules. Instead, correlations between items containing only one rule and WMC were indistinguishable to correlations between items containing up to five rules and WMC (e.g., Unsworth & Engle, 2005). This led to the assumption that not the number of information in a matrix is an important cause for the relationship of WMC and matrix reasoning but, in fact, that the impact of WMC on matrix reasoning is independent of the number of applied rules.

The invariant relationship of WMC and matrix-reasoning across items with a different amount of rules was leading to the assumption that WMC is involved in maintaining solution strategies and responsible for the allocation of resources for rule induction. In particular, studies showed that avoiding distraction (Jarosz & Wiley, 2012), combating proactive interference (Wiley et al., 2011) or successfully storing solution principles over all items (Harrison, Shipstead, & Engle, 2015) in matrix-reasoning tasks was strongly associated with WMC. Notably, it was assumed that these processes are mandatory in *all* items (Embretson, 1995; Unsworth & Engle, 2005).

# 2 Shortcomings in Studies on WMC and Matrix Reasoning

Although it seems that the numbers of rules in an item have no impact on the correlation of matrix reasoning and WMC, this was considered as a premature conclusion as there were both methodological and theoretical shortcomings in the described studies (cf., Little, Lewandowsky, & Craig, 2014).

Besides the consideration of WMC as a composite score or only controlled attention, WMC was mostly assessed by the Ospan, and therefore, by one single task in these studies. This is problematic for two reasons: first, it is questionable whether one task can represent an underlying construct, and second, some authors consider complex span task not as a valid tool to assess WMC (Conway et al., 2005; Conway, Kane, & Engle, 2003).

In addition, the conclusion that the correlation between WMC and matrix-reasoning is invariant across all items was based on point-biserial correlations between WMC and *single* items of matrix reasoning (e.g., Wiley et al., 2011). As some items are solved by 90 percent of participants, and items containing more up to five rules are solved by less than 10 percent, this could bias the correlation and complicates the interpretation, especially since an observation of an invariant relationship is a null effect (Little et al., 2014).

Even more important for the present work is that the matrix-reasoning items in the present study are intermixing different demands on matrix-reasoning. As we described, at least three different processes can be distinguished, which are more demanded in items with multiple rules (storing partial solutions, selective encoding, and goal management). Hence, even when the authors would have found an alteration in the correlation between WMC and the items, it would be hard to conclude on which process this change is due.

# 3 Interim Conclusion 2

WMC can be considered as a pool of different aspects or processes that contribute to a successful maintenance of information for further processing. For the present work, storage capacity, storage and processing, and filtering are considered as essential when investigating the impact of WMC on the three different processes in matrix-reasoning outlined above (storing partial solutions, selective encoding, and goal management). All of the aspects of WMC are uniquely related to gF but direct evidence for the cause of the relationship is scarce. We sought to take the implications of these aspects of WMC into account in the present study in order to disclose why matrix-reasoning items with multiple rules are harder to solve than items with a single rule.

# E The Current Study

The aim of this work was to reconsider the relationships between WMC and matrix reasoning. Especially, we focused on why matrix-reasoning items with multiple rules are more difficult to solve than items with one rule, and how WMC can facilitate the solving of items with multiple rules. As such, we focussed on two main aims.

The first aim was to consider whether higher item difficulties in matrix-reasoning items with multiple rules are due to higher demands in certain processes in matrix reasoning. At this, we focused on the storage of *partial solutions*, *selective encoding* of relevant elements and *goal management*. We designed matrix-reasoning task that allowed the manipulation of the respective processes in an experimental design. This offered the advantage that only the requirements of the respective process could be manipulated, while other influences could be kept constant. In addition, multiple items could be created for a certain process, so that later analyses did not refer to single items only.

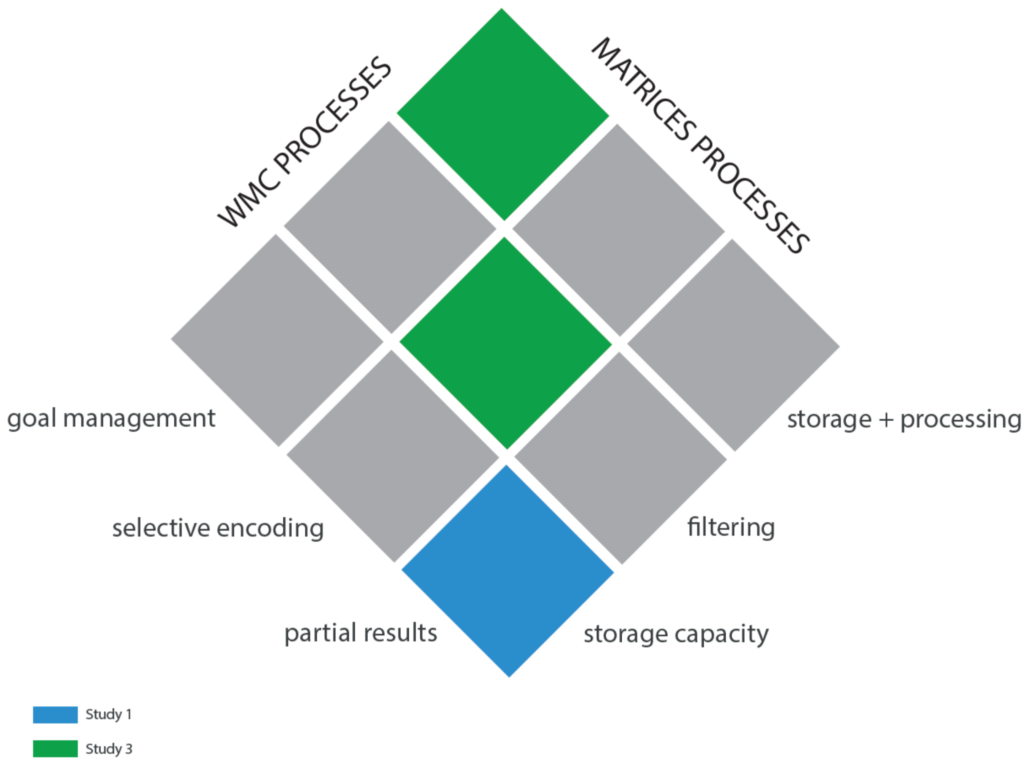



*Figure 17*. Processes of matrix-reasoning and WMC aspects addressed in the current work.

The second aim was to investigate how WMC contributes to these processes. As it has been demonstrated in various studies that the relationship between WMC and gF is driven by different aspects of WMC (e.g., Unsworth et al., 2014), we combined those aspects of WMC with the processes of matrix-reasoning, which are discussed in the literature as plausible aspects for this process. At this, we focused on storage capacity, storage and processing and filtering as aspects of WMC. The assignment of the processes on the different WMC aspects is displayed in *Figure 17*.

**Study 1** focused on whether storing partial solutions is an essential process, which is required in matrix-reasoning items with multiple rules. In addition, we investigated whether storing partial solutions is associated with individual's *storage capacity*, meaning that individuals with higher storage capacity can store more partial solutions and therefore, are more likely to solve items with multiple rules.

Since less is known about the influence of filtering in matrix-reasoning and the operationalization of inter-individual differences in filtering efficiency is rarely investigated, **Study 2** evaluated how filtering can be calculated in order to quantify this relationship to matrix-reasoning and can therefore be regarded as a preliminary study prior to Study 3.

**Study 3** investigated the role of selective encoding in matrix reasoning items. More specifically, the aim was to uncover whether selective encoding demands hampered performance in matrix reasoning and whether *filtering* in working memory facilitated performance in these items. In addition, by means of eye movement analyses, we observed whether selective encoding demands were hampering pairwise comparisons during rule induction as a possible cause for the lower performance since perceptual continuity is reduced in items with multiple rules (cf., Primi, 2002). Study 3 also investigated whether goal management demands hampered performance in addition to selective encoding and whether these demands were associated with *storage and processing.*

In order to address the shortcomings of previous studies on the relation between WMC and matrix-reasoning, we assessed each aspect of WMC with multiple tasks and conducted an SEM approach with latent-variables to ensure a broader construct representation and less biased correlation due to task specificity. In addition, the different matrix-reasoning processes were experimentally manipulated. Hence, for each study, only one process was manipulated and other were held constant. Furthermore, each process-manipulation was performed with *multiple items* in order to avoid analyses that were based on single items.

To manipulate the different demands, purpose-constructed matrix-reasoning tests were developed for each study. For this purpose, both the presentation of the task and the individual items had to be designed so that they optimally represented the manipulated process and largely excluded other confounding factors. Our items were based on the DESIGMA (Becker et al., 2014) as this task allowed for flexible adaption of the design for our needs. The DESIGMA is different in two main aspects to traditional matrix-reasoning tests as the APM. First, the item stem was designed on *a priori* defined construction rules, whereas for the APM only a *post hoc* classification is known. This has the advantage that possible confounding factors can be controlled and new items can easily be created. New items were constructed for each study (except for Study 2), and all items were constructed by applying the six different rules *addition*, *subtraction*, *intersection*, *single element addition*, *completeness*, and *rotation* (see Becker, Schmitz, Falk et al., 2016). A description of the rules and an assignment of which rule is used in which study is shown in the table in Appendix 1.

The second advantage of the DESIGMA was that the solution could not be selected from different response alternatives as in the APM but has to be constructed by single elements given in a “construction kit”. Importantly, studies have demonstrated that constructing the solution instead of selecting, enhances the construct validity of matrix-reasoning tests (Arendasy & Sommer, 2013; Becker, Schmitz, Falk et al., 2016). In addition, this also facilitated the construction of new items as no response alternatives – which are different for each item – had to be constructed.

# F Study 1: Partial Solutions

## 1 Introduction

This study investigated whether the requirement of storing partial solutions contributes to difficulty in items with multiple rules and whether storage capacity is associated with storing partial solutions. At this, we experimentally manipulated storage demands of partial solutions in one matrix reasoning test, which consisted of two versions: the first version provided a sketchpad that enabled externalization of partial solutions and the second version required the storage of partial solutions until the whole item was solved. We hypothesized that enabling externalization of partial solutions to a sketchpad relieves working memory in matrix reasoning. Hence, if storage of partial solutions is necessary for successfully solving matrix-reasoning tests, performance should be better when externalization is provided compared to the condition, in which externalization is prevented (*Hypothesis 1.1*).

More importantly, if storage capacity is involved in storing partial solutions, variability in storage capacity should explain more variance in the matrix-reasoning test in which partial solutions have to be retained compared to the version where partial solutions can be externalized. In other words, the correlation between storage capacity and matrix-reasoning test performance should be stronger in the non-externalized condition compared to the externalized condition (*Hypothesis 1.2*).

Our aim was to experimentally manipulate storage demands in matrix-reasoning tests while controlling for other item characteristics that can influence the solving behavior in matrix-reasoning tests. As there is evidence that the type of rule can affect the solving process matrix reasoning (Embretson, 1998; Green & Kluever, 1992), we counterbalanced the type of rule over all items and across the two conditions of matrix-reasoning tests.

# 2 Methods

## 2.1 Participants and Design

Eighty-five students from Saarland University were tested and received monetary compensation. Due to missing data, one participant had to be excluded from further analyses. The final sample (67% female) had a mean age of 23.3 (SD = 3.8, range 18–35) years.

Participants were assessed in group settings on individual computers with ear protection in order to avoid distraction. The sessions did not exceed two hours. For stimulus presentation and data recording, PsychoPy 1.81.03 (Peirce, 2007) was used.

We applied a within-subject design, in which every participant worked on the matrix-reasoning tasks in both conditions. Additionally, we assessed storage capacity with three types of change detection tasks. Both the order of the three change detection tasks and the order of the two matrix-reasoning tests were counterbalanced across participants. Since a within-subject design was applied, two different item sets for matrix-reasoning test were designed. Each matrix-reasoning test contained one of these two sets. In order to rule out confounding factors of specific items on one matrix-reasoning test, the assignment of item sets to matrix-reasoning tests was also counterbalanced across participants.

## 2.2 Test Methods

### 2.2.1 Change Detection

We measured the individual storage capacity with three different blocks of change detection tasks with a change of color, shape, and orientation (see *Figure 18*). First, a sample array with four or six randomly chosen items was presented for 500 ms. After a blank screen of 900 ms, a test display with only one object appeared at a random position until a response was given.

Participants had to detect if the items' relevant feature (color, shape, orientation) at this position had changed. In 50% of the trials, the test object was identical to the object at the same location in the sample display whereas for change conditions a randomly chosen object that had not been shown within the sample display was presented. Additionally, in the change-conditions, only the critical feature was changing (e.g., color) and the irrelevant feature was fixed (e.g., shape of colored squares). Participants were instructed in advance which feature was potentially changing.

We used eight different colors (green, blue, red, yellow, white, black, violet, cyan), eight shapes and eight orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) of the letter T as stimuli. For each block (color, shape, orientation) 40 trials with set size 4 and 40 trials with set size 6 were

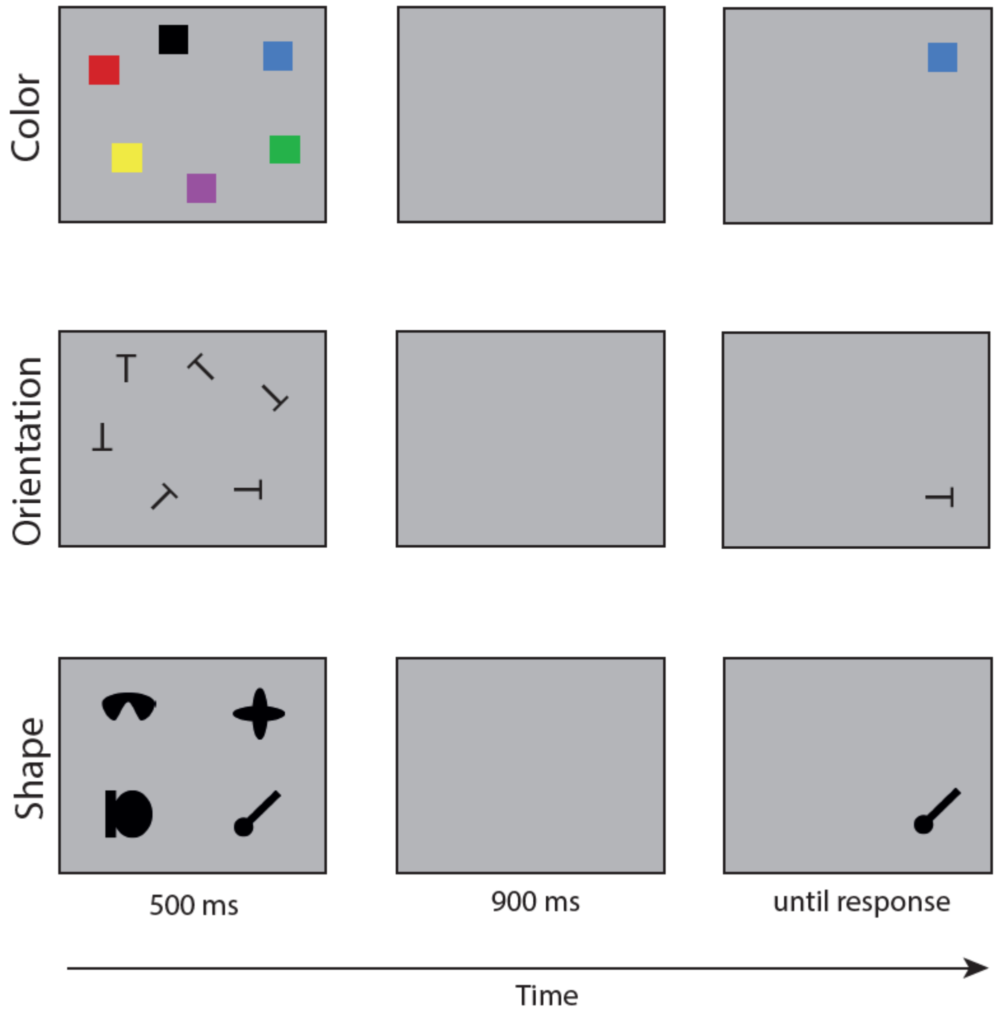



*Figure 18*. Procedure of the change detection task for color, orientation, and shape condition. Figures not drawn to scale. Displayed is a no-change trial for each condition.

presented. Additionally, seven practice trials were shown before each block. All stimuli were presented on an invisible circle around the fixation cross with a radius of 3.92° and were separated by at least 3° (center to center).

### 2.2.2 Matrix Reasoning

Two modified versions of the DESIGMA (Becker et al., 2014) were used. The first version (EXT) allowed for externalization as the opportunity of immediate response while induction was given, whereas the second version (NONEXT) forced subjects to memorize each item's full solution before responding (see *Figure* 19). Each item of EXT consisted of a 3 × 3 matrix with an empty cell (response field) in the right lower corner (see *Figure 19A*). Contrary to traditional matrix-reasoning tests such as Raven's APM, no set of response alternatives including the correct solution was presented. Instead, participants had to construct the solution using 16 elements given in a box below the matrix (construction field). Importantly, these 16 elements were the same elements every matrix was based on. After selecting one element from the construction field, the element appeared within the response field; choosing the same element for the second time deleted this element from the response field. A time limit of 90 s was given for each item, which was empirically determined (see Becker, Schmitz, Göritz et al., 2016). The time remaining to enter the solution was permanently displayed in the upper-right corner of the screen. Additionally, the RESET button offered the opportunity to clear all elements in the response field. Participants were instructed to click on the FINISH button when they believed that they had constructed the correct solution. After 90 s, the item terminated automatically. An optional break of 30 s was given before the next item was displayed. Similar to EXT, items of NONEXT were presented in a 3 × 3 matrix with an empty response field. In contrast to EXT, the construction field in NONEXT remained invisible until the participant confirmed the button NEXT STEP. Thus, participants had to solve the whole matrix mentally before responding and also had to memorize the item's correct solution (inducing step, see *Figure 19B*). After confirming NEXT

STEP, the matrix disappeared, the construction field appeared, and the participants had to choose the elements from the construction field based on their mental representation (responding step, see *Figure 19C*). The solving time within inducing step was 90 s for each item. After 90 s or if participants confirmed NEXT STEP, the responding step started. After the responding step (self-paced), an optional break of 30 s was given before the next item started. Instructions for EXT and NONEXT were held constant, except for the instruction to respond immediately in EXT and to memorize the solution in NONEXT before responding. To familiarize the participants with the testing procedure, one practice item was presented before each version of the matrix-reasoning tests. Both versions of matrix-reasoning tests consisted of 16 randomly presented items. As pointed out in design of the study, two sets of structurally similar items were constructed and were put in either set A or set B.

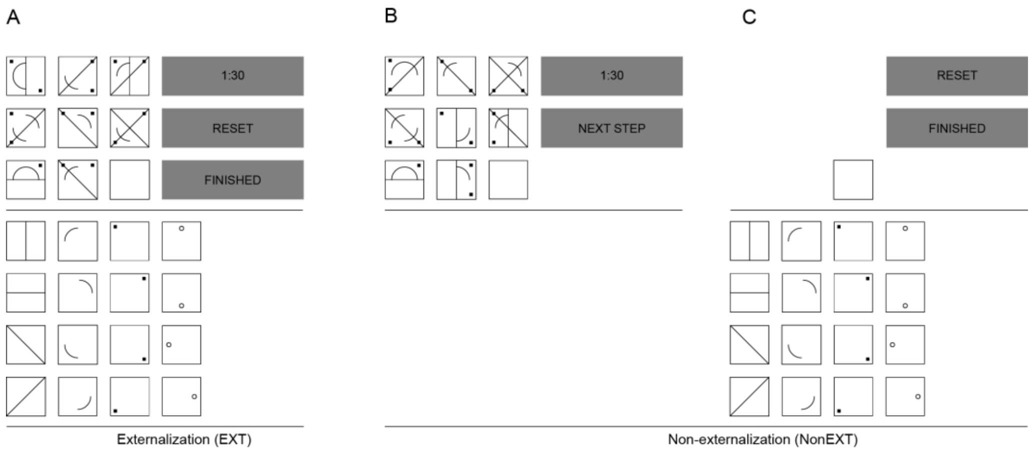



*Figure 19*. Illustration of the two versions of matrix-reasoning (EXT and NONEXT). A. EXT with possibility of immediate response. B. Inducing step of NONEXT. No response can be given. C. Responding step of NONEXT. The matrix disappears directly after confirming next step in B. For better understanding, control button descriptions were translated from German to English for this article.

To hold memory demands based on the number of given visual information during induction constant, each item consisted of three different rules. Combinations of rules and elements were counterbalanced across items. In total, we used the four rules addition, subtraction, intersection, and single element addition (see Appendix 1). Rules were applied over rows. Rules were applied to four possible element groups: lines, circle segments, small squares, and small circles. These elements were exactly the elements shown in the construction field (see *Figure 19A*), ensuring that every item could be solved with elements of the construction field. Within each item, only one rule was only applied to one element group (e.g., addition was applied to lines but not to lines and circles in Item X) so that no rule was repeated within an item. For instance, in *Figure 19A* three rules were applied row-wise to the item: (1) the line elements from the first two cells are summed up in the third cell; (2) the circle elements from the second cell are subtracted from the first cell, and the result appears in the third cell; and (3) the small black squares that are shown in the first or the second cell (but not in both) appear in the third cell.

## 2.3 Data Analysis

### 2.3.1 Change Detection

We estimated K, which is the number of items that an individual can store in working memory as storage capacity (e.g., Cowan, 2001). The estimation is based on the assumption that the participant gives a valid response of each item which is either in memory or otherwise the answer is guessed. Hence, the proportion correct is adjusted for guessing, and therefore, it is an estimate of the proportion of items which were really stored in working memory. We used the standard k-score formula K = set size × (hit rate − false alarm rate) by Cowan (2001) to estimate the individual storage capacity. K represents the number of stored items in memory, set size the number of presented items in the sample display, hit rate the proportion of correctly detected changes, and false alarm rate the proportion of given change responses to non-change trials. We calculated K for each condition and set size individually and used the average of K for set sizes 4 and 6 of each condition for further analyses. Higher values represent higher storage capacity. For correlation analyses, we calculated a joint storage capacity score using the average z-values based on all storage capacity estimates (color, shape, orientation).

### 2.3.2 Matrix reasoning

We used the number of solved rules instead of the number of solved items for each matrix-reasoning test version as an estimate for participants' matrix reasoning ability. This gave us the opportunity to directly observe how many details of the solution were retained for an item. Higher values represent more solved rules. For structural equation modeling, we used item parcels, each consisting of four items that were summed up, resulting in four parcels which served as indicators for the latent cognitive ability factor.

### 2.3.3 Statistical Analyses

Structural equation models and confirmatory factor analysis were conducted using lavaan 0.5–20 (Rosseel, 2012) with maximum likelihood as the estimator. The following conventions were used to assess the global fit of the model: RMSEA < .06, SRMR < .09 and CFI close to .95 (Hu & Bentler, 1999).

# 3 Results

Descriptive statistics and correlations between gF and storage capacity estimates are presented in Table 1. Both matrix-reasoning versions were strongly correlated ($r$ = .81, p < .001).

However, significantly more rules were solved in EXT compared to NONEXT, $t(83) = 2.54$, $p < .05$, $d = 0.28$. Please note that analyses based on the number of solved items revealed a similar result, $t(83) = 2.53$, $p < .05$, $d = 0.28$. Hence, *Hypothesis 1.1* could be confirmed that performance is better when externalization is provided compared to the condition, in which externalization is prevented.

*Table 1.* Descriptives and correlations of Study 1

| | M | SD | Min | Max | Color | Shapes | Ori | Mean K | EXT |
|---|---|---|---|---|---|---|---|---|---|
| Color | 3.27 | 0.7 | 1.3 | 4.45 | - | | | | |
| Shapes | 1.96 | 0.7 | 0.4 | 3.2 | .48 | - | | | |
| Ori | 1.88 | 0.86 | 0.05 | 3.7 | .42 | .37 | - | | |
| Mean k | 0 | 0.78 | -2.07 | 1.36 | .81 | .78 | .76 | - | |
| EXT | 28.94 | 16.17 | 0 | 48 | 0.18 | 0.13 | .25 | .24 | - |
| NONEXT | 26.21 | 15.79 | 0 | 48 | 0.16 | 0.11 | .29 | .24 | .81 |

*Note*: Color K, Shape K and Ori K: estimates for storage capacity for color, shape and orientation condition; Mean K: mean of z-values of Color K; EXT: sum of solved rules for externalize condition; NONEXT: sum of solved rules for non-externalize condition; Correlations based on Pearson-Correlation.

Correlations of single storage capacity tasks with gF ranged from $r$ = .13 to .35 for EXT and from $r$ = .11 to .27 for NONEXT. Importantly, correlations of EXT and NONEXT with storage capacity (Mean $k$) were equivalent, $t(84) < 1$. In order to test differential influences of storage capacity on EXT and NONEXT without measuring error, we used a model with one single storage capacity factor (storage) and one factor for each version of matrix-reasoning (EXT and NONEXT). The resulting model (Model 1, see *Figure 20*) fit the data very well, $\chi2(41) = 41.83$, $p$ = .43, CFI = .999, RMSEA = .016, SRMR = .043. Parameter estimates indicated that storage capacity had a similar relationship to both EXT ($r$ = .28) and NONEXT ($r$ = .27). To test the equality of these correlations, we altered Model 1 by equating the parameter estimates reflecting the correlations between the two matrix-reasoning tests and storage capacity. This restricted model (Model 2) was compared with Model 1. The fitted of Model 2 was excellent, $\chi2(42) = 41.86$, $p$ = .48, CFI = 1.0, RMSEA = 0.0, SRMR = .044. Additionally, there was no significant difference in the global model fit, $\Delta\chi2(1) = 0.03$, $p > 0.05$, which supports the evidence for homogeneous correlations in our bivariate correlation analyses. Hence, we found no evidence supporting *Hypothesis 1.2* that the correlation between storage capacity and matrix-reasoning performance is stronger in the non-externalized condition compared to the externalized condition.

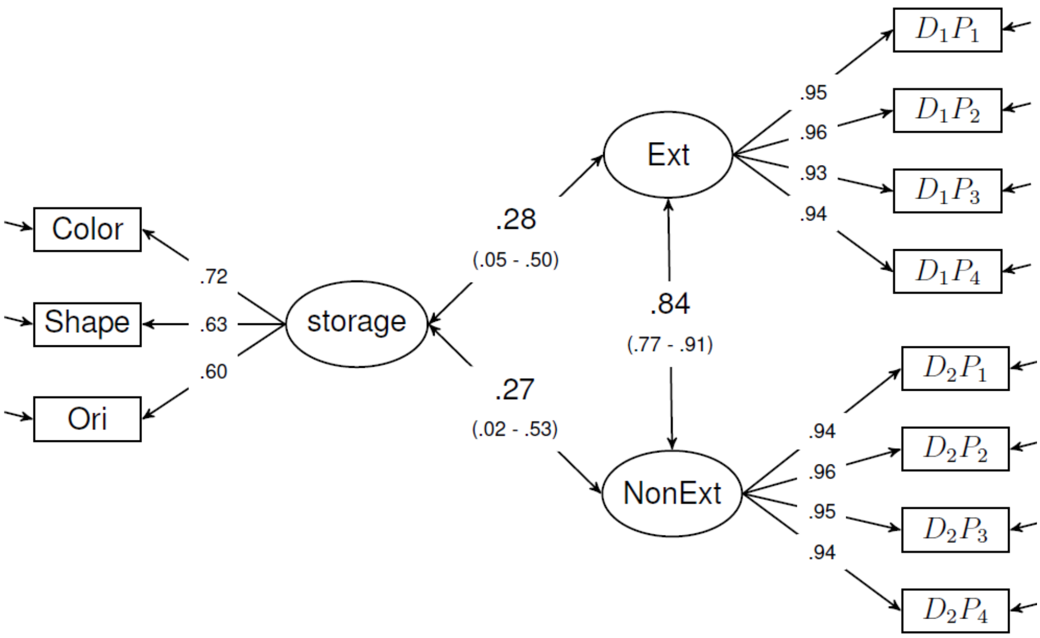



*Figure 20.* SEM for interrelations between storage capacity and the two versions of matrix-reasoning applied in this study (Model 1). Parameters are standardized. CI of standardized estimates are in parentheses. Model fit: $\chi^2(41) = 41.83$, p = .43, CFI = .999, RMSEA = .016, SRMR = .043. *Storage* = storage capacity, *EXT* = externalize condition, *NONEXT* = non-externalize condition, *DnPm* with *n* = matrix-reasoning test index, *m* = parcel index.

# 4 Discussion

The goal of this study was to investigate whether storage of partial solutions is required in matrix-reasoning and whether this is associated with an individual's storage capacity. At this, we experimentally manipulated the demands of storing partial solutions in matrix-reasoning items, which resulted in two matrix-reasoning versions: one versions made it necessary to retain all information until the final response was generated (NONEXT), the other provided the opportunity of using a sketchpad (EXT). We hypothesized that in the EXT condition less storage capacity would be recruited because it was possible for participants to hold partial solutions in the external medium instead of their memory. The NONEXT condition required storage of partial solutions and goals which we expected to increase working memory load. Therefore, we tested whether the correlation between storage capacity and performance in the NONEXT condition would be stronger than between storage capacity and performance in the EXT condition.

Results revealed that performance in the NONEXT condition decreased, compared to the EXT condition, indicating that demands of storing partial solutions hampers performance in matrix reasoning. However, if one considers that this is a small effect (Cohen, 1988), the requirements for storing partial solutions in matrix-reasoning should not be overrated. On average, there would be an advantage of .17 more solved rules in EXT compared to NONEXT, which suggests that storing partial solutions does not strongly influence the solution process.

Notably, the SEM approach revealed that the correlation between storage capacity and performance in matrix-reasoning did not differ between the conditions EXT and NONEXT. This indicates that storage capacity is not related to storing partial solutions in matrix reasoning. In other words, individuals with different storage capacity score on the same level in both matrix reasoning tests, which indicates that the number of stored elements in working memory is not predictive for how many partial solutions can be stored in a matrix-reasoning task.

Taken together, these results challenge the view that storing partial solutions is an essential process in a matrix reasoning, as it was suggested in previous literature about matrix reasoning and related higher-order cognition test (Carpenter et al., 1990; Hitch, 1978; Just & Carpenter, 1992; Mulholland et al., 1980). In addition, this study also failed to provide evidence that storage capacity is related to storing partial solutions although this relation was suspected in recent studies (Unsworth et al., 2014). This is leading to two assumptions: first, storing partial solutions cannot be the predominant requirement leading to higher item difficulties in matrix-reasoning items with multiple rules. Second, storage capacity plays a crucial role in other processes in matrix reasoning or that controlled attention drives the relationship between WMC and matrix-reasoning and not storage capacity.

# G Study 2: Filtering (Preliminary Study for Study 3)

## 1 Introduction

The aim of this preliminary study was to quantify the impact of filtering, assessed by a variant of the change detection paradigm (Vogel et al., 2005), on matrix reasoning. Although several studies emphasized filtering as an important mechanism in working memory (e.g., Liesefeld et al., 2014; Vogel et al., 2005), evidence whether filtering is related to gF or matrix reasoning is scarce. As Study 3 was set out to investigate whether filtering has an impact on a certain matrix-reasoning process, this study observed whether filtering is even related to matrix-reasoning.

Additionally, filtering efficiency is usually estimated by filtering costs, which is a difference score of distractor-absent and distractor-present trials. However, this approach is usually used in studies dealing with mean values. For differential effects in filtering other techniques as the control of common variances of storage capacity and filtering in path analyses or SEM seem also seem promising (cf., Shipstead et al., 2014). This approach is based on the rational that performance on distractor-present trials assess both storage capacity (as information has to be maintained) and filtering (as only relevant information has to be filtered out). When controlling the portion of variance of storage capacity in distractor-present trials (or a latent factor based on these trials), only interindividual variances in filtering remains. This study should evaluate which approach is more appropriate for the differential approach for the subsequent Study 3.

# 2 Methods

## 2.1 Participants and Design

The sample consisted of 105 students from Saarland University. Participants received monetary compensation or course credit. Due to missing data, two participant had to be excluded from further analyses. The final sample (72 female) had a mean age of 23.2 (SD = 5.6, range 17–56) years.

Participants were assessed in group settings (max. five participants) on individual computers with ear protection in order to avoid distraction. The sessions did not exceed two hours. For stimulus presentation and data recording, PsychoPy 1.81.03 (Peirce, 2007) was used.

We applied a within-subject design, in which every participant worked on two matrix-reasoning tests, two storage capacity test, and two filter ability tests. The order of the test was counterbalanced over participants.

## 2.2 Test Methods

### 2.2.1 Matrix reasoning

The DESIGMA (Becker et al., 2014) and the APM (Raven, 1940) were applied to assess matrix reasoning. In order to decrease the durations of the sessions, short versions of the matrix-reasoning tests were applied. For DESIGMA, items with the No. 1-4, 7, 8, 10-17, 19, 22-24, 26, 29-31, 33, 34, 36, and 38 were selected. For APM, items with the No. 1-6, 8-10, 13, 14, 16-18, 21-23, 26, and 29-34 were selected. The items of the short versions were selected based on data of a preliminary study within this project, in which both matrix-reasoning tests were applied in its full length. As such, only these items were selected, which ensure that the over-all difficulty of the test is comparable to the difficulty of the corresponding full length test. Additionally, correlations between the short version and the full length test was .99 for the DESIGMA and .97 for the APM. Since the reliability tends to decrease when a test is shortened (e.g. Schermelleh-Engel & Werner, 2008), we also compared the reliabilities between the short version and the full length test. For both matrix-reasoning tests the reliabilities were comparable (DESIGMA: .95 [full] vs .91 [short]; APM: .87 vs .84).

### 2.2.2 Storage Capacity

As in Study 1, storage capacity was estimated by means of change detection task (Luck & Vogel, 1997). For this study two variants of the change detection task were applied, one with a change of color, and one with a change of orientation. First, a sample array with randomly chosen items was presented for 100 ms, and re-presented after a blank of 900 ms. Participants had to detect if the items' relevant feature (color, orientation) at this position had changed. In 50% of the trials, the sample was identical to the first display, whereas for change conditions a randomly chosen object of the first display was substituted by a random object that was not presented in the first display.

For the color condition, 50 trials with set size 4 and 50 trials with set size 6 were sequentially presented. Additionally, seven practice trials were shown in advance. For the orientation condition 50 trials with set size 3 and 50 trials with set size 5 were sequentially presented. Also, seven practice trials were shown in advance.

For color condition, colored circles with eight different colors (green, blue, red, yellow, white, black, violet, and cyan) were used as stimuli. The circle had a radius of 0.34°. For orientation condition, eight orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) of a black keyhole-like stimuli were used. The size of the stimuli was 0.61° × 1.24°.

All stimuli were presented on an invisible circle around the fixation cross at center of the screen with a radius of 4.49° and were separated by at least 3° (center to center).

### 2.2.3 Filtering

Filtering in working memory was assessed by a modified version of the change detection task in which distractor items were displayed in addition to the targets items (Vogel et al., 2005). We applied two variants of the filter task, one that required to filter out relevant shapes (shape filter), and one that required to filter out relevant colors (filter color; see Appendix 2).

On trials of *shape filter* condition, participants had to *retain the color* of the target object while ignoring the color of the distractor object. Appearance of target items was counterbalanced over participants and could be either squares or rectangles. The distractor object was always the opposite object of the target object. On trials of *color filter* condition, participants had to *retain the orientation* of the target object while ignoring the color of the distractor object. Appearance of target items was counterbalanced over participants and could be either a pink or yellow keyhole-like object. The distractor object was also a keyhole-like object, but always hold in the opposite color as the target object.

The procedure of this task was identical to the change detection that was used to assess storage capacity. Except that participants were additionally instructed to ignore the irrelevant items and only retain the critical features of the relevant items.

For the *shape filter* condition, 100 trials were presented with a random mixture of the two set sizes 4TD0 (4 Targets, 0 Distractors) and 4TD4 (4 Targets, 4 Distractors). Additionally, fourteen practice trials were shown in advance. For the *color filter* condition, 100 trials were presented with a random mixture of the two set sizes 3TD0 (3 Targets, 0 Distractors) and 3TD3 (3 Targets, 3 Distractors). Additionally, fourteen practice trials were shown in advance.

## 2.3 Data Analysis

### 2.3.1 Matrix reasoning

We calculated the number of solved items for both DESIGMA and APM. To receive a joint-matrices-score (*Matrices*), we averaged the number of solved items of DESIGMA and APM. Higher values represent higher matrix reasoning ability.

### 2.3.2 Storage Capacity

Based on the formula of Cowan (2001), we estimated individual K for each set size as in Study 1. Consequently, we received two K-scores for the color condition and two *k*-scores for the orientation condition. To receive a joint-score (*Storage*), all *k*-scores were averaged. The resulting score represents the average number of retained items in working memory.

### 2.3.3 Filtering

The K-score was also used to calculate filtering ability. We estimated K for each set size and condition. Consequently, we received two K-scores for the shape filter and two K-scores for the color filter condition. As previously described, we wanted to observe the impact of filtering on matrix reasoning with two methods. Hence, we calculated two scores for filtering ability. For the first score (*Filter*), the scores of the two conditions *with* distractors (T4D4 for shape and T3D3 for color) were averaged. A higher score represents a higher storage capacity in the face of distracting items.

For the second score (*Filter costs*), the average of the scores *with* distractor was subtracted from the average the scores *without* distractors: mean(T4D0,T3D0) – mean(T4D4,T3D3). A higher score represents higher filtering costs, since the performance is declining when distracting information are presented.

### 2.3.4 Statistical Analyses

Path models were conducted using lavaan 0.5–20 (Rosseel, 2012) with maximum likelihood as the estimator. Initially a structural equation modeling approach was intended to observe the interrelationship between storage capacity, filtering and matrix reasoning. The latent variables were defined by the scores of the single set sizes for storage capacity and filtering, and by parcel for matrix reasoning. However, the model did not converge and therefore, we conducted a path analyses based on the manifest joint-score variables described.

# 3 Results

Descriptives and correlations among the measurement are presented in *Table 2* and *Table 3,* respectively.

*Table 2*. Descriptives of all tasks in Study 2

| | Storage | Filter | Filter costs | Matrices | DESIGMA | APM |
|---|---|---|---|---|---|---|
| Mean | 2.70 | 2.33 | 0.41 | 12.74 | 10.41 | 15.08 |
| SD | 0.55 | 0.47 | 0.35 | 4.90 | 5.92 | 4.67 |
| Min | 1.52 | 1.01 | -0.43 | 2.50 | 1.00 | 2.00 |
| Max | 3.97 | 3.34 | 1.48 | 21.50 | 23.00 | 24.00 |

*Note*: M = Mean, SD = Standard deviation, Min = Minimum, Max = Maximum, Storage = κ as an estimating for storage capacity, Filter = performance on distractor-present trials of the filter task, Filter costs = difference of κ distractor-absent and distractor-present trials, Matrices = joint score of performance in both DESIGMA and APM.

Table 3. Inter-correlations of all tasks in Study 2

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Storage | — | .65 | -.05 | .30 | .18 | .39 |
| 2. Filter | | — | -.54 | .36 | .24 | .45 |
| 3. Filter costs | | | — | -.06 | -.01 | -.12 |
| 4. Matrices | | | | — | .94 | .90 |
| 5. DESIGMA | | | | | — | .71 |
| 6. APM | | | | | | — |

*Note*: Storage = κ as an estimating for storage capacity, Filter = performance on distractor-present trials of the filter task, Filter costs = difference of κ distractor-absent and distractor-present trials, Matrices = joint score of performance in both DESIGMA and APM.

## 3.1 Control of variances

To test the influence of filtering on matrix reasoning when controlling for storage capacity, we conducted a manifest path analyses with *filter* and storage capacity as exogenous variables and matrix reasoning (*Matrices*) as endogenous variable (see *Figure 21*; values left of the slash). Importantly, we regressed *filter* on storage capacity to partial out the variance of storage capacity in *filter*. Consequently, the correlation between *filter* and matrix reasoning is controlled for storage capacity. The analyses revealed that storage capacity is strongly related to *filter* (β = .65, $p < .001$). Additionally, storage capacity is significantly correlated with matrix reasoning (β = .30, $p < .01$). Interestingly, *filter* has a unique significant correlation with matrix reasoning above storage capacity (β = .22, $p < .05$).

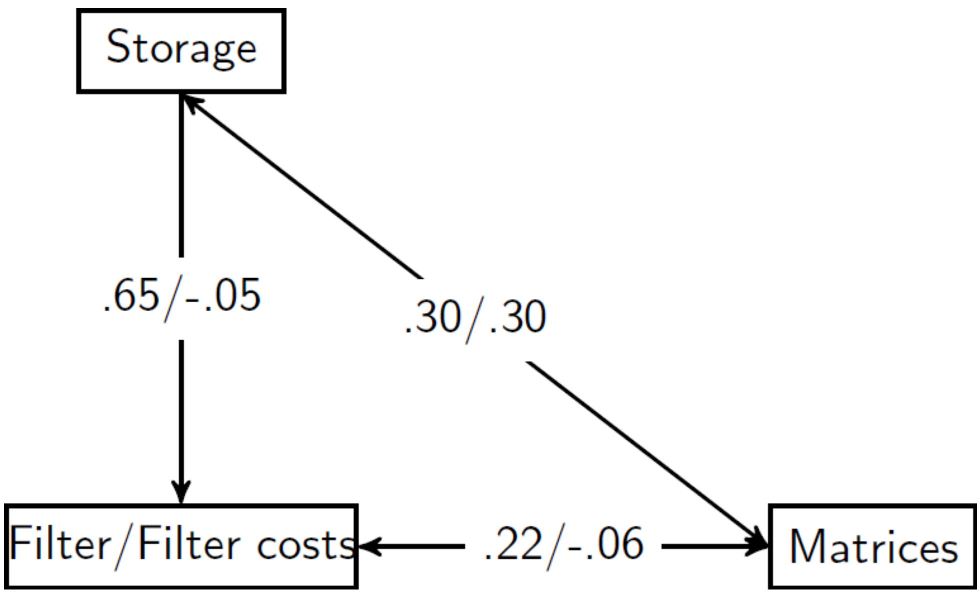



*Figure 21*. Result of path analyses with standardized parameters. Storage = Storage capacity; Matrices = joint matrix-reasoning score of APM and DESIGMA; Filter = performance of the filter tasks in the distractor-present condition; Filter costs = difference-score of distractor-absent minus distractor-present condition.

## 3.2 Filter costs

We also observed the influence of *filter costs* on matrix reasoning besides storage capacity. Therefore, we conducted another path model, in which the influence of *filter costs* and storage capacity on matrix reasoning was observed (see *Figure 21*; values right of the slash). The analysis revealed a non-significant relationship between *filter costs* and storage capacity ($\beta$ = -.05, $p$ = .66). Additionally, storage capacity was positively related to matrix reasoning ($\beta$ = .30, $p$ < .01). However, *filter costs* were not significantly related to Matrices ($\beta$ = -.06, $p$ = .55).

# 4 Discussion

The aim of Study 2 was to quantify the relationship between filtering and matrix-reasoning as research on the differential effects of filtering, assessed by the change detection task, is scarce but we consider this process as highly relevant for selective encoding in matrix reasoning. The results demonstrate that filtering had a significant influence on matrix reasoning. However, this influence was dependent on how filtering efficiency was calculated. The influence of filtering was only significant when the variance of storage capacity was controlled in filtering and the remaining variance was related to matrix reasoning. This approach was based on the assumption that filtering represents both storage capacity and filtering aspects (see Shipstead et al., 2014). However, when subtracting the performance of filter present trials from filter absent trials, which is considered as filter costs, no significant correlation with matrix reasoning was found. Interestingly, filter costs were also not related to storage capacity.

Since several studies could demonstrate an impact of inter-individual differences in filter costs on storage capacity (e.g., Liesefeld et al., 2014; Vogel et al., 2005), we have to ask: why did we fail to show an effect of filter cost on storage capacity or matrix reasoning in this study? One major difference is that we applied a larger sample size than the previous studies. Whereas these studies assessed filtering of around 30 to 40 participant, we assessed filtering of nearly 100 participant. Since there is evidence that correlations only stabilize at larger sample size, and the error of detecting a false positive significant correlation is more likely at smaller sample sizes

(Schönbrodt & Perugini, 2013). However, we are not consider the results of previous studies as fallacious as these studies were aiming at disclosing underlying cognitive mechanisms of filtering and were not focusing on correlation analyses per se. In addition, filter costs were also based on electrophysiological potentials in these studies (e.g., Vogel et al., 2005) and not on the behavioral performance, which could also cause the differences in the results.

In sum, this study demonstrates that controlling storage capacity in filtering when predicting matrix reasoning is more promising than observing the impact of filtering based on filter costs on matrix-reasoning. For this reason, we will take the first approach into account for the subsequent study.

# H Study 3: Selective Encoding and Goal Management

The third study was set out to address these three issues: (1) Are selective encoding demands present in matrix reasoning items with multiple rules besides goal management, (2) Are there indications that pairwise comparisons are hampered in items with selective encoding demands, (3) can filtering in working memory facilitate processing in matrix reasoning items with multiple rules, and (4) is storage and processing associated with goal management?

The first and second issue was addressed in Study 3A. We expected that selective encoding played a crucial role in matrix reasoning items with multiple rules as information of the current processed rule had to be encoded and irrelevant information has to be blocked (Meo et al., 2007; Primi, 2002). In addition, we expected that pairwise comparisons during rule induction were hampered as more time was needed to separate relevant from irrelevant element groups and respondents had to disengage from irrelevant elements groups (Primi, 2002; Meo et al., 2007). Moreover, we expected that goal management was required in addition to selective encoding as the decomposition and serial processing of problems is one of the core assumption of matrix-reasoning processing (Carpenter et al., 1990).

The third and fourth issue was addressed in Study 3B. We expected that *filtering* in working memory facilitated processing in items with selective encoding demands. Furthermore, if *storage and processing* is associated with goal management, successful storage and processing should facilitate item processing when goal management is required in items with multiple rules.

# 1 Study 3A

To independently manipulate selective encoding and goal management demands, three versions of matrix-reasoning were developed, which were adapted from the DESIGMA (Becker et al., 2014). Items of the *first version* contained one rule that had to be solved (One-rule, *1R*, see *Figure 22A*). Therefore, only rule induction was required to infer the correct solution since no irrelevant information was presented and the item had not been composed in problems. In items of the *second version*, two irrelevant rules were applied in addition to a single relevant rule (One rule plus noise, *1RN*, see *Figure 22B*). The participant was informed in advance, which rule was relevant, and which elements had to be ignored. In contrast to 1R, this condition should require selective encoding in addition to rule induction since irrelevant elements had to be blocked. In items of the *third version*, three rules were applied, and the participants were instructed that all rules were relevant for this item (Three rules, *3R*, see *Figure 22C*). Hence, goal management was required in addition to rule induction and selective encoding: The item had to be decomposed in problems, and for successfully solving the item, the rules had to be processed serially (goal management). When inferring the underlying rule of each element group (rule induction), elements from other rules had to be blocked (selective encoding). Consequently, in 3R some elements additionally existed that were *currently* irrelevant when processing the current rule. Hence, we referred to "irrelevant information" for both conditions 1RN and 3R.

We expected that performance was hampered when irrelevant information was presented since selective encoding was required, and further decreased when goal management was required as the respondent had to decompose the problem and has to process the rules serially (*Hypothesis 3A.1*).

The second goal of Study 3A was to find evidence for the causes of a reduction of performance due to selective encoding demands. We used eye movement analyses during solving to obtain indicators if perceptual continuity was disrupted (Primi, 2002). As it requires some time to extract the relevant information from irrelevant information, we hypothesized that irrelevant information in a matrix led to longer dwell times on a matrices cell (*Hypothesis 3A.2*). Additionally, due to the time-consuming selection process, there is less time available for pairwise comparisons that are necessary to detect changes of the visual elements between the cells, which is required to infer an underlying rule. Hence, we expected that the number of pairwise comparisons was lower in items where irrelevant information (1RN, 3R) was presented compared to the condition where only one rule (1R) was presented (*Hypothesis 3A.3*).

## 1.1 Method

### 1.1.1 Participants and design

Forty-three students from Saarland University were tested and received monetary compensation or partial course credit for their participation. Due to invalid eye tracking recordings, five participants had to be excluded from further analyses. The final sample consisted of 38 students (66% female) with a mean age of 21.53 years (SD = 3.24, range 17-36). Participants were assessed in single settings. A within-subject design was applied, in which all participants were solving three conditions of matrix reasoning. Matrix-reasoning items of the three conditions were fully randomly presented.

### 1.1.2 Eye tacking apparatus

For stimulus presentation and data recording PsychoPy 1.83 (Peirce, 2007) was used. Eye movements were recorded by a Tobii TX300 remote eye tracker (Tobii Technologies, 2011). The device consisted of a 23 inche LCD monitor with a resolution of 1920 x 1080 pixels, and an eye tracking module that was placed below the monitor. The sampling rate was set to 60 Hz. Although the eye tracker can compensate for head movements, a headrest was used to ensure a reliable recording of eye movements. Participants were placed 65 cm in from of the monitor, and the light in the laboratory was dimmed.

### 1.1.3 Matrix-reasoning stimuli

Three matrix-reasoning versions (1R, 1RN, 3R) based on the DESIGMA (Becker et al., 2014). For creating the items, the rules *addition*, *subtraction*, *intersection*, *single element addition*, *completeness*, and *rotation* were used (see Appendix 1). Each rule was applied to sets of visual elements (e.g. a small black square or a solid line). To ensure a balanced design, twelve item triplets were designed with one item for each of the three conditions per triplet. A sample triplet is displayed in *Figure 22A-C*. For the condition 1R, the lines in *Figure* 22*A* are governed by the rule *addition*. In the corresponding item for 1RN in *Figure 22B*, the lines are governed by the rule *subtraction.* The rules *overlap* and *rotation* are applied to the black squares or circle segments, respectively. In this item, only the lines are relevant (and cued to the participant in advance), and the other elements are irrelevant. In the corresponding item for 3R in *Figure* 22*C*, the rules *intersection*, *completeness,* and *addition* are applied to the elements lines, black squares and circle segments, respectively.

Both the type of rule and the combination of the type of rule to visual elements was counterbalanced over conditions. Additionally, two structural similar (same rules) but phenotypical different (different operationalization of the rules) versions of item sets were created and randomly assigned to participants.

### 1.1.4 Procedure

All participants were informed in advance which rules are potentially applied in the items as we wanted to observe whether only the perceptual appearance of an item has implications on selective attention and goal management demands independent of the inter-individual differences in pre-existing knowledge about the applied rules. The rules were instructed with example items, which were not used in the main experiment.

At the beginning of each item, a cue was presented for three seconds, which indicated the relevant element group(s) for the current item (*Figure 22D*, cue). Subsequently, a randomly chosen item of one of the three conditions was shown. The presentation time was self-paced but expired after 20 seconds for 1R and 1RN or 60 seconds for 3R. These time limits were empirical determined based on the data of Becker, Schmitz, Göritz et al. (2016) as these data showed the median time of solving an item with one rule was around 20 seconds. Hence, we assumed that 20 seconds are necessary to solve one rule and 60 seconds (3 x 20 seconds) to solve three rules.

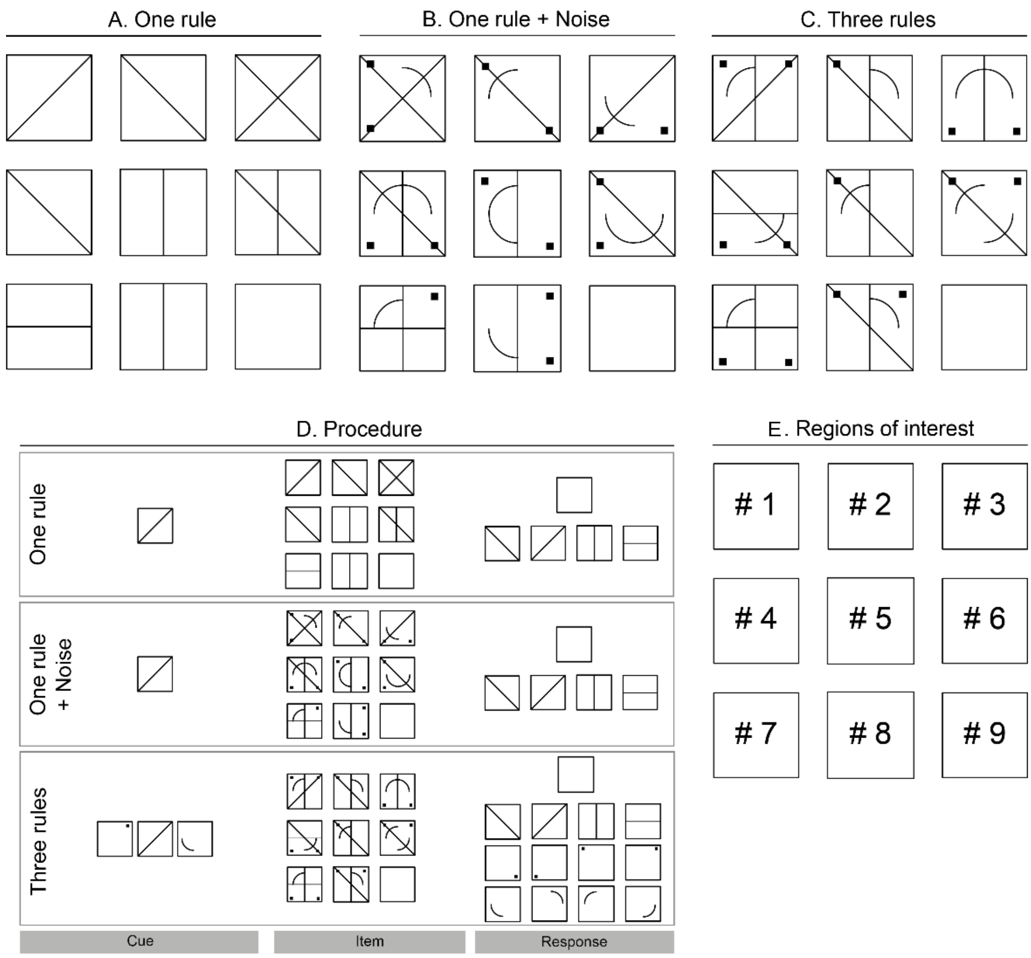



*Figure 22.* Item example of the three different matrix-reasoning versions (A-C); Procedure in Study 3A (D); Regions of interest of the eye tracking analysis (E).

Participants were instructed to solve the item and remember the total response (*Figure 22D*, item). In the next step, the answer could be inserted in the solution field by construction the solution from the construction field (*Figure* 22*D*, response). Additionally, only the relevant element groups that are cued beforehand were shown in the construction field. This is of particular importance for 1RN since selective encoding demands should be induced during rule induction and potential interference with different element groups should be avoided during the response. After each item, a break of five seconds was given.

### 1.1.5 Measures

*Performance in matrix-reasoning*. We defined performance as the number of solved rules relative to the number of applied rules in an item. Consequently, in 1R and 1RN a score of 0 or 1 could be reached for each item, and for 3R the scores 0.33, 0.67 and 1 could be reached for each item. The scores were averaged for each condition resulting in one score for each of the three conditions.

*Eye tracking measures*. Nine regions of interest (ROI) were defined, one for each of the nine matrices cells within the item stem (see *Figure 22E*). Each region of interest had a size of 5.13° x 5.13° with a distance of 3.08° between the regions of interest.

The time on each ROI and toggles between the ROIs were recorded during testing. As successful solving the item requires a row-wise inspection of the matrix from left to right, all gaze data was excluded before the gaze was on the first or third ROI at the first time. Additionally, all gaze data were excluded after the gaze was on the eighth or ninth ROI for the last time for each item. This procedure ensured that we isolated the process of active rule induction from other processes like "orientation" at the beginning or the ending of the item processing (for a similar approach see Hayes, Petrov, & Sederberg, 2011)

For each item, the *mean dwell time* on a cell before shifting to another cell was calculated for each cell and averaged over all items of one condition. The ninth cell was not taken into account as this cell was not containing visual information and therefore, no visual encoding was necessary. This resulted in one mean dwell time for each of the three conditions.

As an estimate for pairwise comparisons, we used the number of *cell toggles*. For that, we counted the number of fixation shifts of one cell to its neighboring cells of the same row and back. For instance, when the gaze was first on cell #1, then on cell #2 (or cell #3) and returned to cell #1, one cell toggle was coded. We calculated a relative score by dividing the cell toggles for each item by the dwell time of the gaze on all matrices cells. Hence, the score can be interpreted as the number of cell toggles per second when the gaze was on the cells (for a similar approach see Vigneau et al., 2006). The number of cell toggles was averaged for each condition resulting in one score of cell toggles for each condition.

## 1.2 Results

### 1.2.1 Performance

The descriptives are summarized in *Table 4*. To test *Hypothesis 3A.1*, a repeated-measurement ANOVA with condition (R1, 1RN, R3) as factor and performance as dependent variable revealed a significant effect, $F(2,74) = 16.45$, $p < 0.001$, $\eta^2 = 0.31$, see *Figure 23A*. Notably, performance was significantly higher in R1 than in R1N, $t(37) = 2.39$, $p = 0.02$ , $d = 0.39$. Additionally, performance of R1N was higher than in R3, $t(37) = 3.48$, $p < 0.01$, $d = 0.57$. Hence, results indicate that performance decreased when selective encoding was required and further declined when goal management was required, which supports *Hypothesis 3A.1*.

*Table 4.* Descriptives of Study 3A

| | Mean | SD | Min | Max |
|---|---|---|---|---|
| Score 1R | 0.62 | 0.18 | 0.00 | 1.00 |
| Score 1RN | 0.57 | 0.18 | 0.00 | 0.83 |
| Score 3R | 0.51 | 0.17 | 0.00 | 0.78 |
| Dwell time 1R | 0.33 | 0.05 | 0.22 | 0.43 |
| Dwell time 1RN | 0.41 | 0.06 | 0.28 | 0.57 |
| Dwell time 3R | 0.40 | 0.05 | 0.31 | 0.53 |
| Toggles 1R | 0.79 | 0.17 | 0.31 | 1.04 |
| Toggles 1RN | 0.65 | 0.13 | 0.27 | 0.84 |
| Toggles 3R | 0.65 | 0.14 | 0.25 | 0.88 |

*Note*: M = Mean, SD = Standard deviation, Min = Minimum, Max = Maximum, R1 = one rule, 1RN = one rule plus irrelevant information (noise), 3R = three rules.

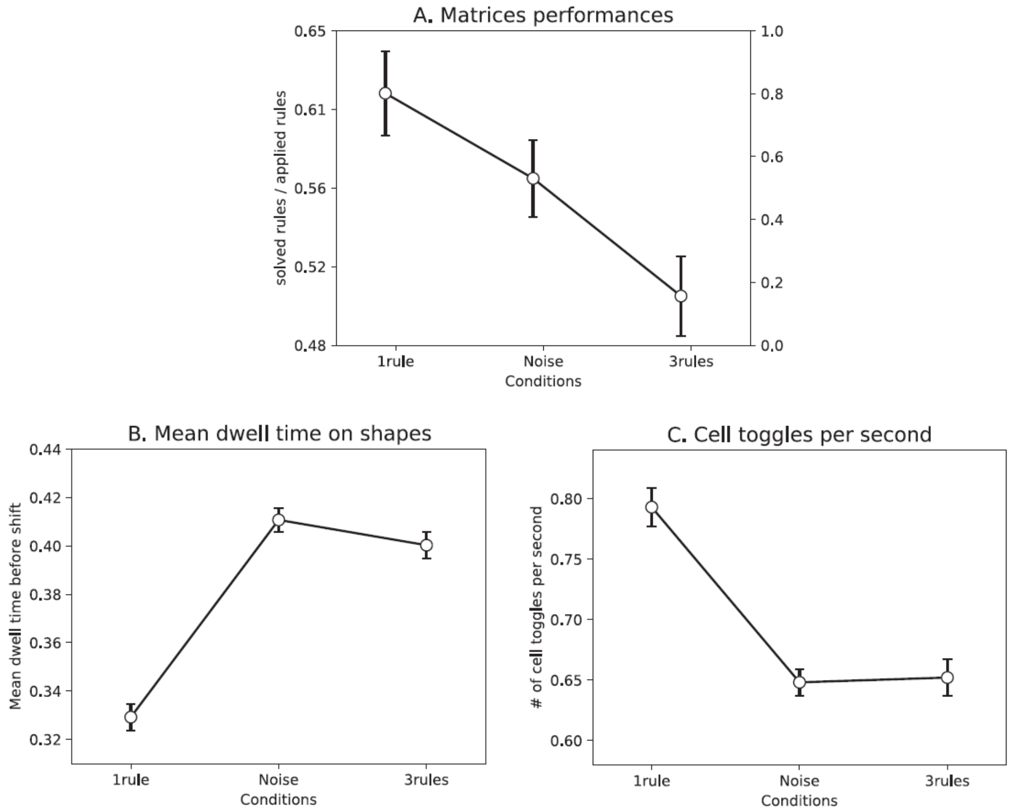



*Figure 23*. Results of Study 3A. Whiskers represent the 95% confidence interval.

### 1.2.2 Dwell Time

To test *Hypothesis 3A.2*, a repeated-measurement ANOVA with condition (R1, R1N, R3) as factor and mean dwell time as dependent variable revealed a significant effect, $F(2,74) = 182.9$ , $p < 0.001$, $\eta^2 = 0.83$, see *Figure 23B*. Furthermore, mean dwell time was shorter in R1 compared to R1N, $t(37) = 18.26$, $p < 0.001$, $d = 2.96$. Hence, this supports *Hypothesis 3A.2* since more time was needed when irrelevant information was present in the items. Additionally, mean dwell time was comparable between the conditions 3R and 1RN, $t(37) = 2.35$, $p = 0.06$, Bonferroni corrected, $d = 0.38$.

### 1.2.3 Cell Toggles

To test *Hypothesis 3A.3*, a repeated-measurement ANOVA with condition (R1, R1N, R3) as factor and cell toggles as dependent variable revealed a significant effect, $F(2,74) = 84.58$ , $p < 0.001$, $\eta^2 = 0.70$, see *Figure 23C*.

Importantly, there were more cell toggles in R1 compared to R1N, $t(37) = 12.33$, $p < 0.001$, $d = 2.0$. Additionally, number of cell toggles were comparable between conditions R1N and R3, $t(37) = -0.37$, $p = 0.72$, $d = -0.06$. Hence, these results support *Hypothesis 3A.3*, since more toggles were performed when irrelevant information was absent compared to the two conditions in which irrelevant information was present.

## 1.3 Discussion

The Study 3A demonstrated that selective encoding demands in matrix reasoning items (1RN) are hampering performance compared to items when only rule induction is required (1R). This is in line with studies that consider selective attention as a source of item difficulty in matrix reasoning (Meo et al., 2007; Primi, 2002), and is the first study to our knowledge that demonstrated that selective encoding demands are present in items with multiple rules. In addition, performance further decreases when goal management is required in addition to selective encoding.

Eye movement data support the finding that there are different demands present in conditions, in which selective encoding is either required or not. Particularly, more time was spent on the cells before shifting to the next cell when selective encoding was required (1RN and 3R) compared to the condition, in which selective encoding was no required (1R). This indicates that the creation of stable representations of the visual material is cumbered by irrelevant information (Primi, 2002) since more time is needed to separate relevant from irrelevant elements. Additionally, respondents showed fewer cell toggles between the cells in the conditions in which irrelevant features are present (1RN, 3R). Pairwise comparisons – operationalized by cell toggles – are a prerequisite for finding similarities and differences to induce an underlying rule (Carpenter et al., 1990; Ragni & Neubert, 2014), and results indicate that irrelevant information in matrix-reasoning leads to less pairwise comparisons.

# 2 Study 3B

The first aim of the second study was to replicate the gradual decrease of performance when irrelevant information and goal management demands are added to the matrix (*Hypothesis 3B.1*). This was done for two main reasons: first, the decrease on performance from 1R to 1RN could simply occur as respondents could have accidentally forgotten the relevant cue in 1RN. Consequently, they could have solved the rule of one of the irrelevant element groups and therefore, did not enter the correct solution of the relevant element group. Second, the final solution had to be maintained before responding since the matrix disappeared in the subsequent response step. Study 1 demonstrated that this demand led to a small decrease in performance in items with three rules, which could contribute to the drop of performance in the 3R condition compared to the 1RN.

The second aim was to inspect whether filtering in working memory was required more in matrix-reasoning items, in which irrelevant information was displayed (1RN, 3R) compared to items in which only one relevant element group (1RN) was presented (*Hypothesis 3B.2*).

The third aim was to identify whether higher goal management demands in matrix reasoning are related to storage and processing. Hence, if goal management is required when multiple rules are applied, storage and processing should be significantly related to the matrix-reasoning versions, in which three rules (3R) are presented but not to versions, in which only one rule (1R, 1RN) is presented (*Hypothesis 3B.3*).

We controlled both filtering and storage and processing for storage capacity to test *Hypotheses 3B.2* and *Hypotheses 3B.3* as this is on the one hand common practice for storage and processing (e.g. Engle et al., 1999) and on the other hand the results of Study 2 revealed that controlling storage capacity in filtering is a promising method to extract filtering ability from the filtering task.

The fourth aim was to observe whether the expected correlation pattern of *Hypotheses 3B.2* (influence of filtering) and *Hypotheses 3B.3* (influence of storage and processing) remained the same when matrix-reasoning was controlled for *g*. Since we did not only wanted to demonstrate what our results tell us about matrix-reasoning processing, but also what our results might indicate for intelligence, we extracted *g* from matrix-reasoning performance. Hence, the remaining variance in matrix-reasoning could be regarded as task-specific variance of matrix-reasoning without *g*. If the correlation pattern of the influence of filtering and storage and processing on matrix-reasoning was remaining the same, we could conclude that this was due to the task-specificity of matrix-reasoning. Otherwise, when the correlation pattern revealed implausible or non-significant results, we could conclude that the correlations of filtering and storage and processing on the matrix-reasoning versions was influenced by an underlying *g*-factor, which would imply that our results are not only relevant for matrix-reasoning but for intelligence in general (*Hypotheses 3B.4*). We employed a screening test of *g* (Kreuzpointner, 2013) as this was an economical test to cover several abilities of *g* in a very broad manner (cf., CHC-theory; McGrew, 2009).

## 2.1 Methods

### 2.1.1 Participants and Design

A total sample of 127 university students (79% females) from Saarland University participated in the study and received monetary compensation or course credit for their participation. The mean age was 22.51 years (SD = 4.16, range 18-44). Participants were assessed in group settings with up to four participants per session. Participants were tested in one session, and sessions did not exceed 2.5 hours.

We applied a within-subject design, in which every participant completed two storage capacity tasks, filter tasks, one screening test of *g*, three matrix-reasoning versions, and three storage plus processing tasks in the described order.

### 2.1.2 Matrix-reasoning

The same matrix-reasoning versions as in Study 3A with the conditions 1R, 1RN and 3R were applied. In contrast to Study 3A, the cued elements and the response were displayed at the same time as the item was presented as we wanted to control for potential confounded memory effects. Additionally, the maximum presentation time of the matrix-reasoning versions was set to 25 seconds for 1R and 1RN, and to 45 seconds for 3R. Furthermore, the rule principles were not explained beforehand to observe whether the results of Study 3A can also be observed under more “common” conditions as rule principles are usually unknown for the respondents.

### 2.1.3 Storage Capacity

*Storage capacity* was assessed by two variants of the change detection paradigm (Luck & Vogel, 1997) for color and orientation. A sample of randomly chosen items (colored squares or tilted Ts) was presented for 500 ms, and participants were instructed to remember the items critical feature (i.e. the color or the orientation of the stimulus). After a blank of 1000 ms, a test display with only one object appeared at a random position until a response was given. In 50 percent of the trials, this object was identical to the object in the first presentation, in the other half of the trials a randomly chosen object that was not shown in the first presentation was presented. Participants had to detect whether the object in the second presentation was identical to the first presentation (no change) or whether the object had changed in the critical feature (change). It is of note that only the critical feature was changed in the change conditions and participants were informed before each block which feature is potentially changing.

Two blocks of change detection tasks were presented, with 24 trials of each set size 3, 4, and 5. Six practice trials were displayed before each block to familiarize the participants with the task.

The first block used colored squares as stimuli and participants were instructed to retain the color of the squares, whereas in the second block the orientations of the letter T had to be retained. The color stimuli had a size of 0.88° x 0.88°, and orientation stimuli had a size of 1.32° x 1.32°. All stimuli had a distance of 2.1° (center-to-center). All stimuli were displayed on an invisible circle with a radius of 4.11°.

### 2.1.4 Filtering

To estimate *filtering*, two blocks of filter tasks were presented. Each block consisted of 24 trials with 3 relevant and 4 irrelevant and 24 trials with 4 relevant and 4 irrelevant items. Four practice trials were displayed before each block to familiarize the participants with the task. In the first block, the participant had to retain the color of squares and ignore the color of the rectangles (see also *Figure 15*). In the second block, participants had to maintain the orientations of the letter T and ignore the orientations of additionally presented bars.

### 2.1.5 Storage and Processing

To assess *storage and processing* (S+P), three shortened versions of complex span tasks (Conway et al., 2005) were used: operation span, symmetry span, and rotation span. All tasks required storing elements from a list while handling competing processing tasks such as solving math operations. List length ranged from 2 to 5 items, which were randomly presented. Each list length occurred once. However, as we only included items in which the accuracy of the processing tasks was above 85 percent, the current list length was repeated when the accuracy was below this threshold to avoid too much missing data. The current list length was maximally repeated two times. Practice trials were displayed before each block to familiarize the participants with the task.

*Operation Span*. The automated operation span (Ospan; Unsworth et al., 2005) consisted of two tasks that were alternately presented. In the first task, participants had to store letters; in the second task, a math operation had to be solved. After all items of the current list were presented, the letters should be chosen in the correct order on a response display.

*Symmetry Span*. The automated symmetry span (SymSpan; e.g. Kane et al., 2004) consisted of two tasks that were alternately presented. In the first task, participants had to store positions in a grid; in the second task, participants had to judge whether a visual pattern picture is symmetric or not. After all items of the current list were presented, the positions should be indicated in the grid in the correct order on a response display.

*Rotation Span*. The automated rotation span (RotSpan; e.g. Foster et al., 2014; Kane et al., 2004) consisted of two tasks that were alternately presented. In the first task, participants had to store orientations of arrows; in the second task, participants had to judge whether a letter or number was presented correctly or mirror-inverted. After all items of the current list were presented, the orientations of the arrows should be indicated in the correct order on a response display.

### 2.1.6 *g*-Screening

A screening of the Leistungsprüfsystem (LPS-2K, Kreuzpointner, 2013) was utilized as a *g*-screening. The test consists of four tests that were based on the subtests 1, 4, 6 and 11 of the Leistungsprüfsystem 2 (LPS-2, Kreuzpointner, Lukesch, & Horn, 2013). The tests were presented block-wise, and a time limit was given for every task.

*LPS 1*. This task required judging whether the orthography of a display word is correct. On every word, one letter was incorrect (e.g. SPAZE; example translated from German to English for better understanding) and participants had to mark the wrong letter (e.g. C).

*LPS 2*. In this task, series of numbers were presented that follow certain underlying rules (e.g. 1,2,3,9,5,6). Participants had to mark the letter that was not following the rule for the current series (e.g. 9).

*LPS 3*. In this task, numbers or letters of the same category were presented in different orientations. In each item, one letter or number was presented mirror-inverted, which then has to be marked by the participant.

*LPS 4*. In this task, a series of number was presented (e.g. 1, 5, 2, 4, 3). Participants had to sum up the numbers and mark the letter that was identical to the last digit of the sum (e.g. 5).

### 2.1.7 Data Analyses

*Matrix reasoning*. For the matrix-reasoning versions, the same scoring was applied as in Study 3A. Consequently, a score of 0 or 1 could be achieved for 1R and 1RN, and a score of 0.33, 0.66 or 1 could be reached for 3R for each item.

*Storage capacity*. For estimating the individual storage capacity, we used the standard formula by Cowan (2001): K = *set size * hit rate – false alarm rate*. We calculated K for each condition (color and shape) and each set size (color 3, shape 3, color 4, …) and averaged the K scores across conditions (see Chow and Conway, (2015) for a similar approach). This resulted in three K scores, one for each set size (S3, S4, and S5).

*Filtering.* The same principle for calculating individual storage capacity was applied for the filter task. For estimating K, only the set sizes of the relevant features were considered (i.e. set size 3 and 4). Since two set sizes were applied, two K-scores for the filter tasks were extracted (FIL 3 and FIL 4).

*Storage plus Processing*. For every complex span task (Ospan, SymSpan, RotSpan), the sum of the number of items that were recalled correctly in their serial position was used as dependent variables.

*LPS-2K*. For each test of the LPS-2K, the number of correct solved items was taken as a score.

*Structural equation modeling*. Structural equation models were conducted using Mplus 7.11 (Muthén & Muthén, 2006) with maximum likelihood as the estimator. The following conventions were used to evaluate the global fit of the model: RMSEA < .06, SRMR < .09 and CFI close to .95 (Hu & Bentler, 1999).

## 2.2 Results

Descriptives of all variables are displayed in *Table 5*. Correlations are displayed in *Appendix 3*.

### 2.2.1 Performance

To test *Hypotheses 3B.1*, we conducted a repeated-measurement ANOVA with condition (R1, R1N, R3) as factor. The analyses revealed a significant effect, $F(2,252) = 93.30$, $p < .001$, $\eta^2 = 0.43$. Simple main effects revealed that performance in R1 was significantly better than in R1N, $t(126) = 4.76$, $p < .001$, $d = 0.42$. Additionally, performance of R1N was better than in R3,

*Table 5*. Descriptives of Study 3B

| | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|
| 1R | 0.58 | 0.20 | 0.00 | 1.00 |
| 1RN | 0.51 | 0.20 | 0.00 | 0.92 |
| 3R | 0.41 | 0.19 | 0.00 | 0.83 |
| S 3 | 2.24 | 0.44 | 0.75 | 3.00 |
| S 4 | 2.46 | 0.61 | 1.17 | 3.83 |
| S 5 | 2.56 | 0.76 | 0.83 | 4.38 |
| FIL 3 | 1.89 | 0.52 | 0.38 | 3.00 |
| FIL 4 | 1.99 | 0.72 | 0.33 | 3.67 |
| Ospan | 12.02 | 2.52 | 2.00 | 14.00 |
| SymSpan | 9.23 | 3.71 | 0.00 | 14.00 |
| RotSpan | 9.04 | 3.29 | 0.00 | 14.00 |
| LPS1 | 39.01 | 10.98 | 0.00 | 58.00 |
| LPS2 | 22.09 | 4.52 | 0.00 | 34.00 |
| LPS3 | 24.35 | 7.33 | 7.00 | 40.00 |
| LPS4 | 19.50 | 6.76 | 1.00 | 41.00 |

*Note*: M = Mean, SD = Standard deviation, Min = Minimum, Max = Maximum, R1 = one rule, 1RN = one rule plus irrelevant information (noise), 3R = three rules, Ospan = Operation Span, SymSpan = Symmetry Span, RotSpan = Rotation Span, LPS = Leistungsprüfsystem.

$t(126) = 8.91$, $p < .001$, $d = 0.79$. Results show that performance declined when selective encoding demands were present and further declined when goal management was required (see *Figure 24*, solid line).

However, two influences lead to a biased result of the contrast between 1RN and 3R. First, we applied different time limits to matrix-reasoning tests for 1R/1RN compared to the 3R. Second, respondents can reach a score of "1" in 1R and 1RN when one rule is solved in. However, in 3R, this score can only be reached when all three rules are solved, which could artificially decrease the performance. When one rule was solved in 3R, the respondent receives a score of ".33". To rule out these two influences, we used the number of solved rules divided by the time the respondent needed to process the items as dependent variable (see *Figure 24*, dashed line). Consequently, the score can be interpreted as "number of solved rules per second". Although the ANOVA revealed a significant main effect for

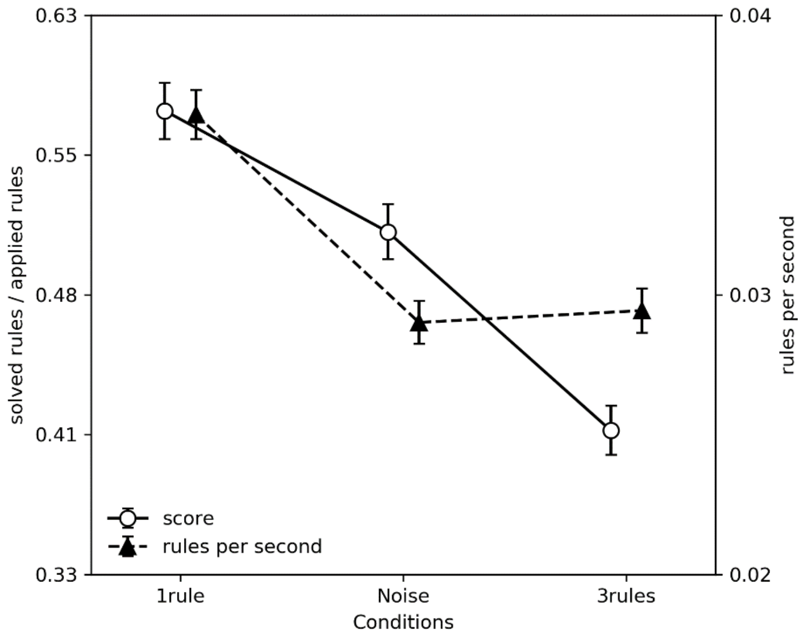



*Figure 24*. Results of Study 3B. Whiskers represent the 95% within-subject confidence interval.

condition, $F(2,252) = 67.61$, $p < .001$, $\eta^2 = 0.35$, the contrasts showed that there was only a decrease in performance from 1R to one 1RN, $t(126) = 10.1$, $p < .001$, $d = 0.9$, but performance remained stable between 1RN and 3R, $t(126) = -0.65$ $p = .52$, $d = -0.06$. Hence, *Hypothesis 3B.1* could be partially confirmed: performance declines when selective encoding demands are added compared to items with only a single rule, and further declines when goal management demands are added. However, the latter effect strongly depends on the calculation of the dependent variable.

### 2.2.2 WMC and Matrix Reasoning

To test *Hypotheses 3B.2 and 3B.3*, a structural equation model was conducted, in which the relations of both *filtering* and *storage and processing* (S+P) on the three versions of matrix-reasoning tests were observed (see *Figure 25*). As such, we controlled both *filtering* and *storage and processing* for the individual storage capacity to reveal the unique impact of both aspects of controlled attention on matrix-reasoning since both tasks also assess storage capacity besides controlled attention as previously described (e.g., Shipstead et al., 2014). To this end, we defined a factor “storage capacity” by the three storage capacity estimates. In addition, both the estimates of filtering and storage and processing loaded on the same storage capacity factor to ensure that storage capacity was controlled in these estimates. Furthermore, the filtering estimates loaded on a separate *filtering* factor and the complex span task loaded on a separate *storage and processing* factor. Consequently, these factors represent the variances of the estimates, which are not based on differences in the individual storage capacity but in differences in filtering or storage and processing. The model revealed an excellent global fit, $\chi^2(29) = 27.75$, $p = .53$, CFI = 1.000, RMSEA = .000, SRMR = .031.

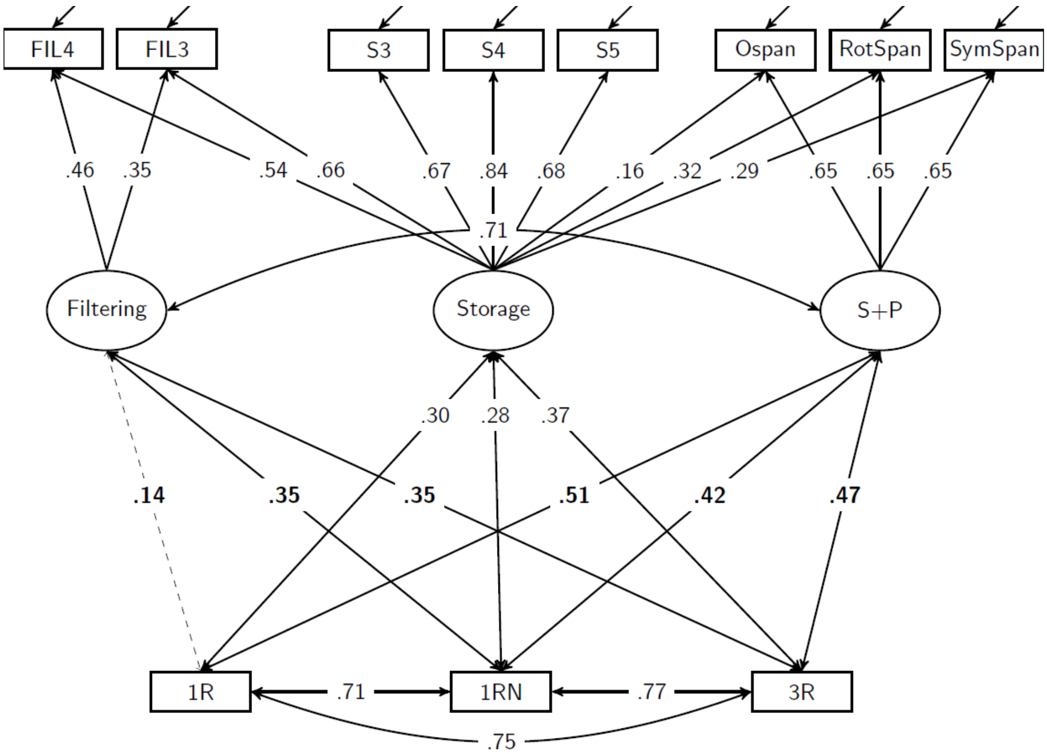



*Figure 25.* Structural equation model for Study 3B with standardized parameters. Storage = Storage capacity, S+P = Storage and processing

Inspection of latent correlations revealed that *filtering* had no significant relation to 1R ($r = .14$, $p = .41$), but shares a significant amount of variance with both matrix-reasoning versions containing multiple information ($r = .40$, $r = .37$ for 1RN and 3R, respectively; all $p < .05$). In contrast, *storage and processing* had a similar influence on all three matrix-reasoning versions ($r = .51$, $r = .42$, $r = .47$ for 1R, 1RN and 3R, respectively; all $p < .001$). Therefore, results revealed evidence for *Hypotheses 3B.2* that filtering is required in items with irrelevant elements (1RN) or in items in which elements are temporitly irrelevant when processing a current rule (3R). However, *Hypotheses 3B.3* could not be confirmed as storage plus processing was not more associated with items requiring *goal management* (3R) compared to items with no goal management demands (1R, 1RN).

### 2.2.3 WMC and Matrix Reasoning without *g*

To test whether the results of *Hypothesis 3B.2* and *3B.3* are equivalent, when controlling for *g* in matrix-reasoning, we regressed *g* on all three matrix-reasoning versions. All three versions were strongly related to a latent *g*-factor based on the four tests of the LPS-2K. The data fitted the model very well, $\chi^2(69) = 70.81$, $p = .41$, CFI = .998, RMSEA = .014, SRMR = .042. A regression of the three scores on *g* revealed that *g* explained 33.1, 30.3, 36.0 percent of variance in 1R, 1RN and 3R, respectively.

Hence, there were indications that all three matrix-reasoning tests shared a substantial amount with *g*. To demonstrate that the substantial correlations of working memory with the matrix-reasoning tests conditions is based on the *shared* variance of the matrix-reasoning score with *g*, we conducted a similar model as in Fig 5 with the modification of regressing the three matrix-reasoning scores on *g.* The fit of the model was excellent ($\chi^2(72) = 75.02$, $p = .38$, CFI = .996, RMSEA = .018, SRMR = .043). In this model, both *filtering* ($r = -.24$, $r = .05$, $r = .00$ with 1R, 1RN, and 3R, respectively; all correlations $p > .05$) and *storage and processing* ($r = .14$, $r = .06$, $r = .08$ with 1R, 1RN, and 3R, respectively; all correlations $p > .05$) were no longer significantly correlated with the residual variance of the three matrix-reasoning tests scores. Hence, *Hypotheses 3B.4* could be confirmed by showing that both the substantial correlations between *filtering* and 1RN, 3R and of SP to all matrix-reasoning tests scores are associated with the part of matrix-reasoning tests scores that share variance with *g*.

## 2.3 Discussion

First, results demonstrate that selective encoding demands are hampering performance, which replicates the results of Study 3A. Additionally, it also could be replicated that goal management demands are hampering performance in addition to irrelevant information. However, this result strongly depends how performance was calculated, and whether different time limits and scorings for the items in 1R and 1RN compared to 3R were take into account. Due to our design, we cannot clearly discriminate how these two time limits affect the solution process.

Furthermore, results demonstrate that performance in matrix-reasoning items with irrelevant information was associated with more efficient filtering in working memory. Hence, individuals with better abilities to filter out relevant information in working memory were more able to solve matrix reasoning items with multiple information, which is in line with previous assumption of the crucial role of selective encoding in matrix-reasoning and its relation to WMC (Meo et al., 2007; Primi, 2002). Since this is dependent on the level of individuals' *g*, it can be concluded that ignoring irrelevant features when encoding the matrix, is not only a basic perceptual processing in matrix reasoning but, in fact, can be related to intelligence in general.

However, storage and processing is not related to a greater extent to the matrix-reasoning version that requires goal management. This indicates that storage and processing as one aspect of WMC is not related to goal management in matrix reasoning. Since goal management is assumed to be the most essential process in matrix-reasoning (e.g., Carpenter et al., 1990), our results are in contrast with this view. Moreover, our findings suggest that the difficulty in an item is determined by the requirements for selective encoding and goal monitoring plays a less important role.

# I General Discussion

## 1 Overview

This study offers new evidence for the question why matrix-reasoning items with multiple rules are more difficult to solve than items with one rule. Models on matrix-reasoning consider the *storage of partial solutions*, *selective encoding* of relevant elements in the matrix, and the keeping track of the solution process, which is known as *goal management*, as important processes that are demanded when multiple rules are applied (Carpenter et al., 1990; Mulholland et al., 1980; Primi, 2002). Notably, all of these processes are theoretically related to WMC and especially goal management is considered as the essential process (Carpenter et al., 1990). However, evidence, whether these processes are actually demanded in items with multiple rules is scarce and the interplay with WMC is still an ongoing question.

In the present work, we experimentally manipulated the respective processes in matrix reasoning and observed whether these processes are essential to solve an item with multiple rules. In addition, we observed how specific aspects of WMC contribute to a successful solving. The results have implications on our understanding of matrix reasoning processing but also on the involvement of WMC in matrix-reasoning and gF, which we want to discuss in the present chapter.

## 2 Implications on Matrix Reasoning

### 2.1 Partial Solutions

The first study investigated whether storing partial solutions is a significant process for matrix-reasoning items with multiple rules. In addition, it was examined whether this process is related to the individual storage capacity. We hypothesized that individuals with a higher storage capacity could maintain more partial solutions of items with multiple rules, and therefore, reach a higher score. This assumption was based on studies indicating that items with multiple rules require the storage of intermediate products or partial solutions in gF-related tasks like matrix reasoning (e.g., Unsworth et al., 2014). All of these studies consider WMC as the bottleneck in storing partial solutions or indicate that each partial solutions require an individual slot in working memory (e.g. Mulholland et al., 1980).

Contrary to this assumption, we could not provide evidence that storing partial solutions is an essential process, which is demanded when multiple rules are applied. We used a design with two versions of matrix-reasoning tests: one version with the possibility to externalize the partial solutions and one version without this possibility. Although performance was significantly better in the externalization condition compared to the non-externalization condition, this was only a small effect as less than a fifth of a rule was solved more, on average, than in the non-externalization version. More specifically, if storage of partial solutions would be *the* essential process to solve items with multiple rules, one would expect that performance in the externalization version would be on the same level as items with only a single rule as in these items no partial solutions of multiple rules have to be stored. However, this was not the case. We calculated the mean item difficulty of the externalized condition, which was lower ($p \sim .40$) as for one rule in a comparable test ($p \sim .70$, Becker et al., 2016) indicating that other demands than the storage of partial solutions leading to a higher item difficulty.

Besides the weak effect of storing partial solutions on the performance level, the invariant relationship between storage capacity and the two matrix-reasoning versions is especially challenging the view that storing partial solutions is an essential process related to WMC. As storage capacity describes how much distinct information can be maintained in working memory, this result means that the potential to maintain more information in working memory does not contribute to the successful solution of an item requiring the storage of partial results. Hence, individuals with a higher storage capacity have no advantage over individuals with less storage capacity in solving items, in which the storage of partial results is required.

Although this finding remains in contrast with the literature highlighting the need of storing partial solutions (e.g., Mulholland et al., 1980; Just & Carpenter, 1992), others also have argued that storing partial solutions might not be the essential process in matrix reasoning. For instance, Verguts and DeBoeck (2002) argued that “just storing partial solutions” (p. 39) cannot be the fundamental underlying mechanism that drives the relationship between working memory and matrix-reasoning processing. This indicates that other processes are relevant for solving matrix-reasoning items, and therefore, other processes are demanded when solving items with multiple rules.

However, is the storage of partial solutions, in fact, not a relevant mechanism in matrix reasoning? We believe the answer is both yes and no. Since we interpret a null-effect in this study, it is difficult to assume that the storage of partial solutions does not play any role. Especially, since we could find a weak effect on the performance level, which indicates that higher demands in the non-externalization condition hamper the finding of a successful solution to some degree.

What supports the "yes" is that it is questionable whether these increased requirements represent the storage of partial solutions. In fact, other causes for the lower performance in the non-externalization condition are possible, which are not associated with a higher demand for storing partial solutions. For instance, visual operations during generation of the response figure as visual operations performed during generation of the response figure (e.g., visual search or element encoding) can cause interference (e.g., Oberauer, Farrell, Jarrold, & Lewandowsky, 2016). That means, for instance, that all information is successfully stored until the whole matrix is solved and during the response phase the presented elements in the construction field interfere with the mental representation of the partial solutions, which causes that the solutions of all rules cannot be recalled properly. This could also indicate that not only the storage capacity aspect of WMC but controlled attention as another aspect of WMC is related to these requirements. Hence, blocking interference during the response generation and drawing attention to the maintained partial solutions could be an essential mechanism that is related to this requirement. However, as this study was designed to investigate whether the potential of storing more information in working memory (storage capacity) is associated with storing partial solutions, and as we did not assess controlled attention, we can only speculate about this involvement.

What supports the "no" is that due to our experimental design it is difficult to generalize our findings to all matrix-reasoning tests and to gF. Our design is based on a subtraction logic (e.g. see Donders cited in Sternberg, 1969), which implies that adding or removing storage demands ideally does not change other processes, for example, the induction process of solving a reasoning problem. This is an assumption, and its validity cannot be directly proven. Therefore, further studies have to disentangle the storage demands caused by other processes from the demands of storing partial solutions, and how both are moderated by the response format.

This would address the second limitation, as well. In order to manipulate the storage demands of partial solutions, we assessed reasoning abilities by the DESIGMA, which has a different test format to conventional matrix-reasoning tests. Conventional matrix-reasoning tests present the item stem together with response options. Although conventional tests do not allow externalizing responses of partial solutions as in the externalize condition of the present study, the displayed response options with correct partial solutions may provide some memory support. Additionally, these displays can be used for a guided search for transformations of additional features, which would reduce the storage demands during the solution process. The results of this study are therefore only first data on the contribution of storing partial solutions to performances in conventional matrix-reasoning tests. In future studies, conventional matrix-reasoning tests like the APM should be applied in addition to the DESIGMA to allow inferences to matrix reasoning in general. Additionally, the inclusion of alternative tests to assess gF would ensure the evaluation of the influence of storage demands not only on matrix reasoning but also on gF.

In addition, in both matrix-reasoning versions in the present study, all items contained three rules as we wanted to control for confounding factors based on the amount of given visual information. Typically, items with a broader range of item difficulties are used in matrix-reasoning, and it could occur that the storage of *three* partial solutions is not sufficiently working memory demanding, which could contribute to the weak effect size and the invariant relationship between storage capacity and the two versions. However, we could exclude ceiling effects in performance, which could have been indicating that the task with three rules was too simple for the participants (28.94 of 42 solve rules for externalization and 26.21 of 42 solves rules for non-externalization condition) indicating that items are differentiating in an average range. Items with more rules could potentially lead to floor effects, which could bias the analyses. Nonetheless, further studies should apply matrix-reasoning tests items with a broader range of item difficulty (i.e., a broader range of the number of rules).

In sum, the first study indicated that the demands of storing partial solutions were not related to storage capacity. Since this association is assumed from the perspective of gF (e.g., Mulholland et al., 1980) and from working memory research (e.g., Unsworth et al., 2014), this study makes a significant contribution to clarify the requirements of storage of partial solutions and WMC, although we have found a null effect. Further studies need to replicate whether storing partial solutions does in fact not play any role in matrix reasoning. With regard to the overarching construct gF, it must also be investigated whether the conclusions of this study can easily be applied to other gF tasks since the importance of storing partial solutions can vary between tasks. For example, storing intermediate steps in mathematical operations seems to be more important than merely maintaining parts of a solution matrix-reasoning test (cf. Hitch, 1978).

## 2.2 Selective Encoding

Another process, which is more demanded in matrix-reasoning items with multiple rules compared to items with single rules, is selective encoding. Whereas the storage of partial solutions can be considered as a more “passive” process in matrix-reasoning as only information has to be maintained until the complete item was solved, selective encoding is involved in rule induction, and therefore at the very core mechanism of matrix-reasoning. As the encoded information is directly serving as input for the rule induction (Carpenter et al., 1990), the encoding of the right information for the current rule is essential for finding the correct solution. Irrelevant information distract the respondent from finding the underlying rule principle as it disrupts the perceptual continuity during the pairwise comparison (Meo et al., 2007; Primi, 2002), which is essential for the induction process (see Carpenter et al., 1990; Spearman, 1927).

The third study set out to investigate how irrelevant information in a matrix affects the solving behavior. We focused on three aspects: First, does the performance decrease when irrelevant information is added? Second, can we find an indicator that the perceptual continuity is disrupted and pairwise comparisons are hampered? Third, does efficient filtering of relevant information in working memory contribute to a better performance in items with irrelevant information?

The novelty of this study was that we did not artificially add irrelevant features such as colors or shading to relevant elements in the matrix as in previous studies (e.g., Primi, 2002). However, we aimed at disclosing whether selective encoding demands are present in conventional items with multiple rules, which are overlapping. At this, we applied three matrix-reasoning versions: one version with one rule and two versions with three rules. One specialty of one of the versions with three rules was that the respondent was asked to solve only one underlying rule of one element group, which was cued beforehand. Hence, this matrix-reasoning version required the solving of one rule plus an additional selective encoding demand. This selective encoding demand was also present in the normal three-rule condition but the three-rule condition also required other processes as goal management. By contrasting the three versions of matrix reasoning, it was possible to observe the impact of selective encoding in matrix reasoning with multiple rules.

The results demonstrated that items with selective encoding demands are harder to solve than items with one rule but easier to solve than items with three rules. This indicates that selective encoding demands are hampering performance compared to items, in which only rule induction is required. It also indicates that three rules are also harder to solve as other demands, such as goal management, place an even heavier burden on the solution process.

We also could show that the average time on each matrix cell was longer in items with selective encoding demands compared to items with one rules and that less pairwise comparisons are performed in the selective-encoding items. We take these findings as an indicator that the solution process is hampered due to the time needed for the segmentation of the whole figure in single parts and the disengagement from irrelevant information. As a consequence of this time-consuming process, less time is available to perform the pairwise comparisons. This finding is in line with research claiming that selective encoding demands in matrix-reasoning items hamper the perceptual continuity (Primi, 2002). It also finds support in working memory research, that demonstrated that attentional capture from irrelevant information is a time-consuming process and varies between individuals with different ability levels (Fukuda & Vogel, 2009, 2011).

In support, we demonstrated that individuals with higher filtering ability in working memory receive higher scores in matrix-reasoning items with selective encoding demands. Interestingly, the substantial relationship between filtering and matrix-reasoning performance was not significant for the one-rule version and did not differ between the selective-encoding and the three-rule condition. We take this as evidence that selective encoding is an essential process in items with multiple rules, and that this process is associated with filtering in working memory as one important aspect of controlled attention.

Additionally, we could demonstrate that the influence of filtering on matrix reasoning with multiple rules is not due to task-specificity in matrix reasoning tests but can be related to intelligence. As our results showed that the correlation-pattern was implausible or non-significant when we removed *g* from the matrix-reasoning tests, we concluded that selective encoding and its relationship to filtering is an essential mechanism in intelligence, which is also supported in the literature. On the one hand, the importance of selective encoding in intelligence was already highlighted by Raven or Sternberg as they argued that reasoning is a process to “make meaning out of confusion” (e.g. Raven, 2002, p.2) or the decision for what is “relevant or irrelevant” (Sternberg, 1986, p. 284). On the other hand, filtering in working memory is already considered as an essential mechanism for higher-order cognition (e.g., Fukuda & Vogel, 2009; Vogel et al., 2005). However, further studies have to replicate this finding and demonstrate whether selective encoding demands is an essential mechanism in gF, especially as we have applied a screening test of *g* in the present study, which was only based on four sub-tests (Kreuzpointner, 2013).

In addition, studies have to investigate the role of filtering in traditional matrix-reasoning items with multiple rules as the APM since there are different demands on *correspondence finding*. In items of the matrix-reasoning test in the present work, all element groups can be easily separated, and on each element group, only one rule is applied. For instance, the *circle* elements can be clearly distinguished from the *lines* and the *squares* in the left example in *Figure 26*. Hence, it is evident for the solver that the item likely consists of three rules and that every element group is governed by one rule. However, in the right example of *Figure 26*, which is an item from the APM, it is harder for the solver to indicate on first sight, which elements have to be separated, and how many rules are potentially applied. The reason is that some rule or elements are merged together into one single shape (e.g., see Preckel, 2003), which could require a different processing method than solving an item from the present work. In fact, the solver first has to infer which elements could be governed by one rule and which elements by another rule, which is described as correspondence finding (Carpenter et al., 1990). Our study cannot resolve whether this process requires the same filtering mechanism, which is involved in matrix-reasoning items in the present work. However, being aware of this difference between our matrix-reasoning tests and others as the APM, we do not consider this as a major limitation, as only some items of the APM are constructed like the one in the right example of *Figure 26*, and the most items are similarly constructed as the matrix-reasoning tests in our study.

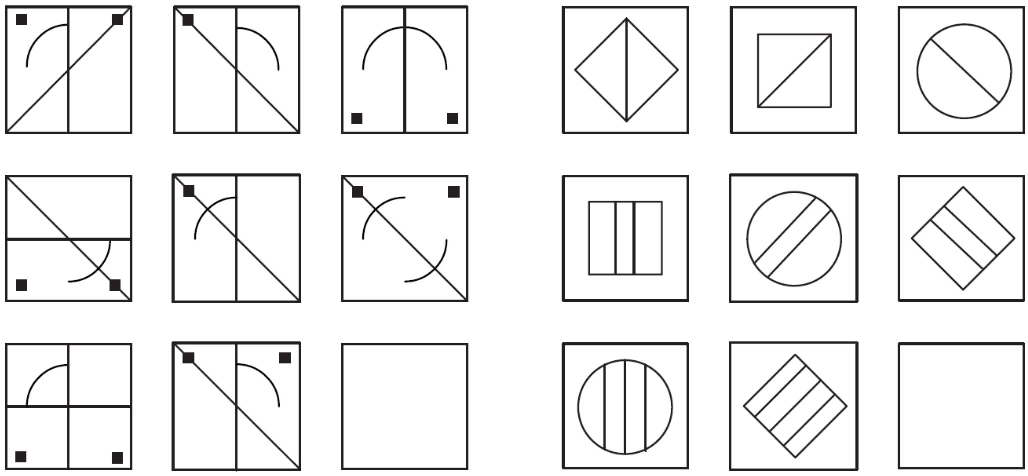


*Figure 26*. Illustration of an item applied in the present study (left, adapted from the DESIMGA; Becker et al. 2014) and items with more difficult correspondence finding (right, adapted from the APM; Raven, 1940).

In summary, a substantial amount of difficulty in matrix-reasoning items with multiple rules is driven by selective encoding demands as overlapping element groups have to be segmented and only the relevant information for the current rule have to be selectively encoded. Notably, this process is facilitated by filtering in working memory. This is a striking result as a previous literature suggested that difficulty in an item with multiple rules is determined by goal management demands, and to our knowledge, this is the first study demonstrating that some parts of the item difficulty associated with goal management can be attributed to selective encoding demands.

## 2.3 Goal Management

To observe whether goal-management demands hamper performance in addition to selective encoding demands and whether these demands are also associated with WMC, we also observed the effect of goal management in the third study. Since three rules were applied to items with selective encoding and to items goal-management demands, the visual appearance was identical in both versions. The only difference was that all items have to be solved in items with goal-management demands whereas only one rule had to be solved in items with selective encoding demands. Hence, items of the goal-management condition required the same processing as items in the selective-encoding condition plus goal management.

We found that performance further decreased when goal-management demands were added compared to the selective encoding condition. This can be taken as an indicator that goal-monitor demands require additional resources in addition to selective encoding demands and that this has an influence on the performance. However, as already outlined in Study 3, there are some aspects, as the scoring and time limits, in the experimental design, which make an unconditional interpretation of the results difficult. Since analyses taking these influences into account revealed that there was no decrease in performance, this result alerts us to take the influence of goal management carefully in matrix-reasoning items with multiple rules. In support, the influence of *storage and processing* was invariant to matrix-reasoning versions requiring rule induction, selective encoding or goal management. If goal management was associated with storage and processing, we would have expected that storage and processing was related to items with goal management demands (three rules) and not (or to a significantly smaller extent) to items with only rule induction or selective encoding demands.

Although we failed to provide evidence for existence of goal management requirements in matrix reasoning, this is less surprising as several previous studies could also not show an involvement of the goal monitor in matrix-reasoning (Embretson, 1995; Unsworth & Engle, 2005). The idea of goal management was mainly based on computer models but evidence demonstrating the existence of this process was only demonstrated in two studies to our knowledge. Carpenter et al. (1990) reported a substantial correlation between the APM and the Tower of Hanoi, which also requires the building and monitoring of goals. However, it has been seen critical whether this correlation, in fact, is evidence for the presence of goal management demands (Oberauer, Süß, Wilhelm, & Sander, 2007). Loesche et al. (2015) considered the fact that the correlation between a WMC composite score and matrix-reasoning increased when the rules were trained in advance as an indicator that goal management is an essential process. However, other mechanisms (e.g., better LTM support in trained items) could also cause the higher relationship and this can only be considered as indirect evidence.

In fact, our study replicated a finding, which was shown in several studies on matrix reasoning and WMC: goal management demands in items with multiple rules are not requiring more WMC (Salthouse, 1993; Unsworth & Engle, 2005; Wiley et al., 2011). However, with our design we can be more specific: goal-management demands cannot clearly be observed on the performance level and they are not related to *storage and processing*. In addition, we have ruled out some methodological shortcomings that existed in previous studies, as described in the introduction (see also Little et al., 2014).

However, to address a shortcoming of the study, an explanation for finding no evidence of a relation between goal management and WMC could also be an inappropriate operationalization of the specific aspect of WMC. We considered storage and processing as a promising aspect of WMC associated with goal management as both require the storage of intermediate steps while performing a secondary task (e.g., math operation in WMC task and rule induction in matrix reasoning) and a redirection the stored information after the task is completed. However, other aspects of WMC could be associated with goal management, which are assessing *goal management* in a more appropriate way. It is hard to conclude which aspect this could be and we do not want to speculate as the description of goal management and the involvement of WMC is quite vague (e.g. Carpenter et al., 1990). What we want to conclude is that further studies on goal management have to provide a more comprehensive description of goal management on a cognitive level and have to predict which aspects of WMC could be involved.

# 3 Implications on WMC

Although the focus of the present work was on matrix-reasoning processing and its connection to WMC, the results have also implications on the understanding of WMC and its relationship to matrix-reasoning and gF.

## 3.1 Storage Capacity and Storage + Processing

Previous research indicated that *storage capacity* is involved in storing information, which are required for the solving process, like hypotheses, goals and partial solutions (Unsworth et al., 2014). However, direct evidence demonstrating whether storage capacity is involved in storing this kind of information is scarce. In contrast, in the present work, requirements of some of this information were experimental manipulated and the influence of storage capacity on these influences was observed. In the first study, the storage of partial solutions was manipulated, and in the third study the number of rules in a matrix.

The results have shown that storage capacity had a similar influence on matrix-reasoning, independent of the manipulation of the requirements in the items. We already discussed that the influence of storage capacity was not moderated by the requirement of storing partial solutions. In addition, the third study revealed that also the number of applied rules has no influence on the correlation between storage capacity and matrix reasoning. As more rules should require the storage of more hypotheses or goals, this can be taken as evidence that inter-individual differences in storage capacity are not associated with the storage of this kind of information in matrix reasoning.

The same result was also found for storage and processing as assessed by complex span tasks. The third study revealed an invariant relationship between storage and processing and all matrix-reasoning versions indicating that also storage and processing is required in all matrix-reasoning items independent of the item's characteristics. But why are both aspects of WMC, storage capacity and storage and processing, related to all matrix reasoning to the same extent? We want to focus on three different explanations, which are controversially debated in the recent literature.

The *first* explanation is that both aspects of WMC could measure the same underlying mechanism: controlled attention. Although we discussed that there is a clear distinction between storage capacity and controlled attention as both can be considered as different perspectives on the model by Cowan (1995), studies could demonstrate that the two constructs are more similar than one might expect (Shipstead et al., 2012; Shipstead et al., 2014). The large amount of variance, which these both include lead to the assumption "that storage capacity performance is not strictly driven by a limited-capacity storage system (e.g., the focus of attention; Cowan, 2001), but may also rely on control processes such as selective attention and controlled memory search" (Shipstead et al., 2012, p. 608). This is supported by studies demonstrating the strong relationship between storage capacity and filtering as one specific aspect of controlled attention (Cowan & Morey, 2006; Vogel et al., 2005). From a theoretical point of view, it is also evident that storage capacity is more than just storage. Based on the model by Cowan (e.g., 1995) storage capacity describes the size of the focus of attention, and therefore, how many items can be maintained above a certain threshold of activation. As previously described, items in the focus of attention are in an interference-free and highly accessible state, which already implies that controlled attention is needed to bring the information in the focus of attention and therefore, is also assessed by the storage capacity tasks as the change detection.

Given the fact that both aspects of WMC could rely on the same mechanisms, which is controlled attention, the invariant relationship of these two aspects of WMC on matrix-reasoning can be can be assigned to previous literature, which also demonstrated an invariant relationship between controlled attention and matrix-reasoning (e.g., Unsworth & Engle, 2005; Wiley et al., 2011). These studies concluded that controlled attention is necessary for *all* items in blocking interference and distraction. For instance, controlled attention is associated with preventing learned rules, from previous items, to interfere with the induction process of the current rule, which might be similar to the previously learned rule but different in some aspect (Wiley et al., 2011). In contrast, controlled attention is also required in the opposite case when pre-learned rules have to be recalled and applied to the current item, in which the same rule is applied (Harrison et al., 2015).

This specific role of controlled attention in matrix reasoning is also leading to the *second* explanation for the invariant relationship between the two aspects of WMC and all matrix-reasoning items. Several studies have shown that *secondary memory retrieval* is associated with both WMC and gF (Shipstead et al., 2014; Unsworth & Spillers, 2010). In essence, secondary memory retrieval describes the controlled retrieval of information from LTM, which includes the generation of retrieval cues and monitoring of the retrieval process (Unsworth et al., 2014; Unsworth & Spillers, 2010). In terms of matrix reasoning, this means that individuals with high secondary memory retrieval ability could better retrieve hypotheses or rule principles, which are no longer within the focus of attention (Unsworth et al., 2014). More specifically, learned *rule principles* from pre-learned items could be more efficiently retrieved from secondary memory by individuals with higher secondary memory retrieval ability, and it can be monitored whether these principles are appropriate or not. Thus, secondary memory retrieval can be considered as an essential mechanism for *rule induction*, although further studies have to provide evidence to support this assumption.

Notably, several studies demonstrated that both aspect of WMC, storage capacity, and storage and processing, are related to secondary memory retrieval (Shipstead et al., 2014; Unsworth et al., 2014; Unsworth & Engle, 2007), which indicates that both aspects of WMC are related to the same underlying mechanism. Hence, in terms of the present study, this suggests that both aspects of WMC have an invariant correlation to all matrix-reasoning items since secondary memory retrieval is required in the rule-induction process, which is required in all items.

A *third* explanation is the building and breaking of bindings, which has been shown to be associated with both WMC and gF. This idea is based on the considerations by Oberauer and colleagues (e.g., Oberauer, 2002; Oberauer et al., 2007). The authors describe a slightly modified version of the embedded-process model by Cowan (e.g., 1995). They also suggest that working memory is activated LTM. However, they assume that the focus of attention can only maintain one item on which transformations can be performed for further processes. The focus of attention is embedded in a region of direct access, which can maintain the usual three to four elements, also described by Cowan. In terms of gF, they describe that WMC based on this model is the ability to build and break arbitrary bindings between the elements within the region of direct access and the one element within the focus of attention (Oberauer et al., 2007). More specifically, for matrix reasoning, this indicates that WMC is involved in maintaining a current representation of the problem in the focus of attention and bind it with other representations of the problem or pre-learned rule principles in the region of direct access to form a new rule or transformation. Hence, this process is also required for rule induction and therefore, for all matrix-reasoning items. Storage capacity is especially associated theoretically with building and breaking bindings when looking for an explanation for the correlation to gF (e.g., Chuderski & Nęcka, 2012), but also storage and processing is assumed to be involved in building and breaking bindings (Wilhelm, Hildebrandt, & Oberauer, 2013). Hence, storage capacity and storage and processing could be related to gF since both are rely on the same process (building bindings), and as this process is required in all items for rule induction, this could explain why the relationship of both storage capacity and storage and processing is invariant for all matrix-reasoning items.

## 3.2 Filtering

However, besides the invariant relationship of storage capacity and storage and processing on all matrix-reasoning items, this study also highlights a particular aspect: We could demonstrate in the second study that filtering in working memory was explaining unique variance in matrix reasoning above storage capacity. More importantly, the relationship between filtering and matrix reasoning was moderated by the demands on selective encoding. We already described that this has an essential implication about our understanding of matrix-reasoning processing. However, it also has implications for underlying mechanisms in working memory.

First, to our knowledge, this is the first study, which demonstrates the substantial role of filtering in one gF-related test. Although filtering is particularly important in working memory research, no study has so far demonstrated a *significant* involvement of filtering in gF (e.g., see Shipstead et al., 2014).

Second, this demonstrates that the operationalization of WMC tasks is essential to uncover an influence on certain characteristics in matrix reasoning. If we had captured filtering with a controlled attention task that covers other processes besides selective encoding, as is the case in complex span tasks, we would not have been able to show this effect. In other words, independent of the specific question posed in the current study, the effect shows that it is necessary to clarify the requirements in the gF-test that is to be connected to WMC and to specify WMC according to these requirements. Hence, our work illustrates that WMC should not be seen as a unitary construct under the *label* WMC, but rather it must specify, which processes can have a *functional* relationship to gF. In a superordinate picture, this is in line with the WMC literature, which describes WMC as a multi-faceted construct, which assumes different processes in WMC, and that different processes have different contributions to gF (e.g., Unsworth et al., 2014).

To sum up, our results have two main implications on WMC: first, it demonstrates that *storage capacity* and *storage and processing* are related to matrix-reasoning independent of the requirements on storing partial solutions, selective encoding or goal management, which indicates that these aspects of WMC are involved in all items. The cause for this invariant relationship could be that they rely on similar mechanisms (controlled attention, secondary memory retrieval, and building bindings), which are mandatory in all items requiring rule induction. However, further evidence has to be provided in future studies to support these assumptions. Second, to describe the relationship between WMC and matrix reasoning on a functional level, WMC tasks have to be utilized, which cover in particular the specific process one wants to uncover in matrix reasoning.

# 4 A Revised Process-Model of Matrix Reasoning

By integrating the conclusions of the last two sections, we want to present a revised process-model of matrix reasoning (see Figure 27). The traditional process-model by Carpenter et al. (1990) posits two main processes: rule induction and goal management. Especially, goal management is assumed to be associated with WMC and mainly responsible why items with multiple rules are harder to solve than items with a single rule, which require no goal management. However, as previously discussed, the results of the present study challenge this view.

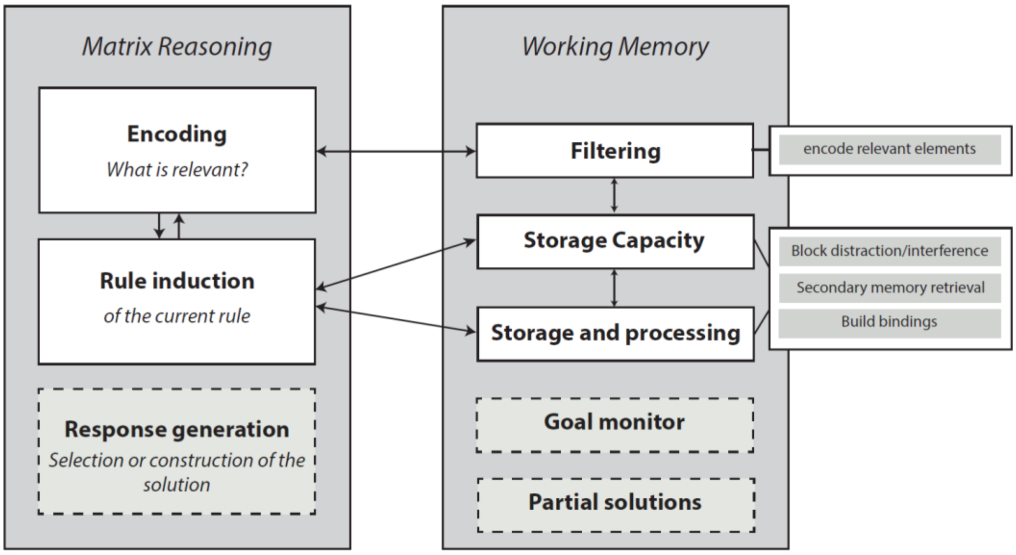



*Figure 27*. Revised process-model of matrix reasoning

The revised model is based on the original model by Carpenter et al. (1990, see also *Figure 8)*. The matrix-reasoning processes are illustrated in the left column and working memory in the right column. However, in contrast to Carpenter et al. (1990), WMC is not considered as one construct but consists of several *aspects* of WMC, which is in line with the multi-faceted view of WMC (Unsworth et al., 2014). Another difference is that we make clear predictions how aspects of WMC in this model interact with matrix-reasoning processes, whereas Carpenter et al. (1990) are quite vague in their description about the involvement of WMC in matrix reasoning (except for the goal monitor).

Based on our model, we assume that information first has to be encoded. In items with one rule, only relevant information has to be encoded and no filtering is required. In items with multiple rules, however, the respondent has to segment the element groups and has to encode selectively only relevant information, which serves as input for the rule induction process. Essentially, without efficient *filtering* in working memory the rule cannot be induced correctly (e.g., due to hampered pairwise comparisons), which leads to a poor performance. This is also one main difference to the original model, which describes the encoding process as relevant but does not consider it as source inter-individual differences. In addition, as outlined above, some matrix-reasoning tests from the APM require the identification which element groups are governed by a rule (correspondence finding), which we suspect to be easier in the applied matrix-reasoning items in the current study. To facilitate correspondence finding we suggest that rule induction and encoding do influence each other mutually. This relationship was also indicated by Mulholland et al. (1980) in geometric analogies, in the sense that the rule induction process can give hints as to which elements might be relevant for the problem.

After the elements of the current rule are encoded, the rule is induced, which is associated with *storage capacity* and *storage and processing* as the invariant correlation between these aspects on WMC and matrix-reasoning suggest an involvement in rule induction. Based on the literature, several processes such as blocking interference or distraction, retrieval of information from secondary memory, or the building of bindings between elements in working memory seem to be involved during rule induction.

Since we found a null effect for storing partial solutions and goal management, we can only speculate as to how they are involved in the revised model. What we can say based on our results, is that both are not sources of inter-individual differences or are not associated with inter-individual differences in WMC, respectively. However, especially for partial solutions, we believe that this process is required in matrix reasoning, as the information has to be stored in some medium, but that this is equally performed by all respondents. This argument is in line with the considerations by Embretson (1983, 1995, 1998) as she argues that some processes in cognitive tasks are relevant but are not a source of inter-individual differences and therefore, equally performed by all individuals.

After all rules were induced, the response can be generated by *selecting* (as in APM) or *constructing* the answer (as in DESIGMA). However, the consideration of the response generation process is beyond the scope of the present work and is only displayed for the sake of completeness.

# 5 Conclusion

With the aim of explaining why items with multiple rules in matrix-reasoning are harder to solve than items with one rule, the present study has examined the processes of storing partial solutions, selective encoding and goal management in matrix-reasoning tests and investigated possible impacts of WMC on these processes. Particularly, based on the present study, great importance can be attributed to the influence of selective encoding on the solution process and its connection to filtering in working memory. Storing partial solutions and goal management, on the other hand, seem to play a less important role than originally assumed in the literature. The study shows important implications for our understanding of matrix reasoning and its connection to WMC, and we hope to have provided a significant piece of a large puzzle why people are more intelligent than others.

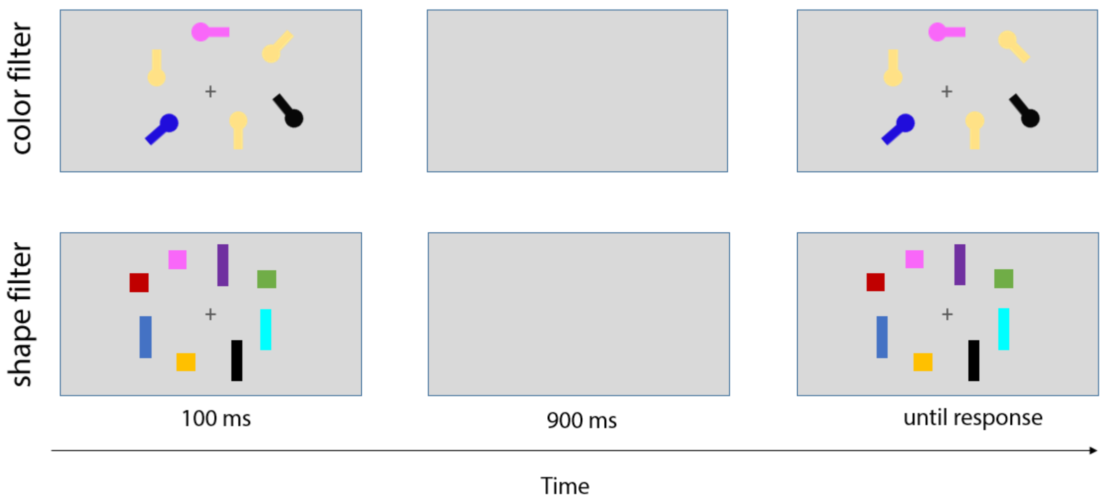
# J Appendix

*Appendix 1. Description of the applied rule in the present work.*

| Rule type | Description | Applied in |
|---|---|---|
| *Addition* | Elements of first two cells are summed up in third cell. | All studies |
| *Subtraction* | Elements of second cell are subtracted from elements in first cell and the result is shown in third cell. | All studies |
| *Intersection* | Elements that are shown in first and second cell are presented in third cell. | All studies |
| *Single element addition* | Elements that are shown in first or second cell are presented in third cell. | All studies |
| *Completeness* | The same set of elements is presented in every row of the matrix. | Study 2 and 3 |
| *Rotation* | Elements are rotated across the cells. | Study 2 and 3 |

*Note*: Description based on Becker, Schmitz, Falk et al. (2016)

*Appendix 2. Description of the applied rule in the present work.*



*Appendix 3. Intercorrelations of Study 3B*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. 1R | — | .71 | .75 | .18 | .27 | .19 | .28 | .21 | .37 | .47 | .39 | .18 | .50 | .37 | .37 |
| 2. 1RN | | — | .77 | .26 | .23 | .10 | .34 | .30 | .31 | .41 | .32 | .15 | .50 | .32 | .30 |
| 3. 3R | | | — | .30 | .32 | .15 | .37 | .35 | .38 | .48 | .34 | .23 | .53 | .30 | .42 |
| 4. S 3 | | | | — | .56 | .43 | .48 | .38 | .17 | .21 | .17 | -.05 | .17 | .07 | .09 |
| 5. S 4 | | | | | — | .57 | .55 | .42 | .10 | .29 | .26 | .07 | .28 | .11 | .24 |
| 6. S 5 | | | | | | — | .43 | .43 | .09 | .17 | .22 | .03 | .23 | .12 | .08 |
| 7. FIL 3 | | | | | | | — | .52 | .31 | .35 | .32 | .22 | .43 | .14 | .26 |
| 8. FIL 4 | | | | | | | | — | .30 | .39 | .38 | .22 | .38 | .24 | .25 |
| 9. Ospan | | | | | | | | | — | .44 | .52 | .35 | .38 | .23 | .31 |
| 10. Sym | | | | | | | | | | — | .51 | .33 | .50 | .36 | .36 |
| 11. Rot | | | | | | | | | | | — | .32 | .45 | .37 | .23 |
| 12. LPS1 | | | | | | | | | | | | — | .44 | .35 | .23 |
| 13. LPS2 | | | | | | | | | | | | | — | .48 | .50 |
| 14. LPS3 | | | | | | | | | | | | | | — | .29 |
| 15. LPS4 | | | | | | | | | | | | | | | — |

*Note*: R1 = one rule; 1RN = one rule plus irrelevant information (noise); 3R = three rules; S 3, S 4, S 5 = K for change detection with set size 3, 4 or 5; FIL 3, FIL 4 = K for distractor-present trials in the filter task with 3 or 4 distractors; Ospan = Operation Span; Sym = Symmetry Span; Rot = Rotation Span; LPS = Leistungsprüfsystem.

# K References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. https://doi.org/10.1037/0033-2909.131.1.30

Arendasy, M. E., & Sommer, M. (2013). Reducing response elimination strategies enhances the construct validity of figural matrices. *Intelligence*, *41*(4), 234–243. https://doi.org/10.1016/j.intell.2013.03.006

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G.A. Bower (Ed.), *Recent Advances in Learning and Motivation* (pp. 47–90). New York: Academic Press.

Becker, N., Preckel, F., Karbach, J., Raffel, N., & Spinath, F. M. (2014). Die Matrizenkonstruktionsaufgabe. *Diagnostica.* Advance online publication. https://doi.org/10.1026/0012-1924/a000111

Becker, N., Schmitz, F., Falk, A., Feldbrügge, J., Recktenwald, D., Wilhelm, O.,. . . Spinath, F. M. (2016). Preventing Response Elimination Strategies Improves the Convergent Validity of Figural Matrices. *Journal of Intelligence*, *4*(1), 2. https://doi.org/10.3390/jintelligence4010002

Becker, N., Schmitz, F., Göritz, A., & Spinath, F. M. (2016). Sometimes More Is Better, and Sometimes Less Is Better: Task Complexity Moderates the Response Time Accuracy Correlation. *Journal of Intelligence*, *4*(3), 11. https://doi.org/10.3390/jintelligence4030011

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence*, *8*(3), 205–238. https://doi.org/10.1016/0160-2896(84)90009-6

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning=working memory≠attention. *Intelligence*, *33*(3), 251–272. https://doi.org/10.1016/j.intell.2005.01.002

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven

Progressive Matrices Test. *Psychological Review*, *97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*: Cambridge University Press.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Chow, M., & Conway, A. R. A. (2015). The scope and control of attention: Sources of variance in working memory capacity. *Memory & Cognition*, *43*(3), 325–339. https://doi.org/10.3758/s13421-014-0496-9

Chuderski, A., & Nęcka, E. (2012). The contribution of working memory to fluid reasoning: Capacity, control, or both? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1689–1710. https://doi.org/10.1037/a0028465

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*(3), 278–295. https://doi.org/10.1016/j.intell.2012.02.010

Cohen, J. D. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Colom, R., Flores-Mendoza, C., Quiroga, M. Á., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences*, *39*(5), 1005–1014. https://doi.org/10.1016/j.paid.2005.03.020

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition*, *34*(1), 158–171.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Cowan, N. (1988). Evolving conceptions of memory storage selective attention and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163.

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition*, *34*(8), 1754–1768. https://doi.org/10.3758/BF03195936

Cowan, N., & Morey, C. C. (2006). Visual working memory depends on attentional filtering. *Trends in Cognitive Sciences*, *10*(4), 139–141.

Donders, F. C. (1969). Over de snelheid van psychische processen.: Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868-1869, Tweede reeks, II, 92-120. *Attention and Performance 11, Acta Psychol.*, *30*, 412–431.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span, *93*(1), 179–197.

Embretson, S. E. (1995). The role of working memory capacity and general control processes in intelligence. *Intelligence*, *20*(2), 169–189. https://doi.org/10.1016/0160-2896(95)90031-4

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*, *3*(3), 380–396.

Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach. *Journal of Experimental Psychology. General*, *12*(3), 309–331.

Evans, T. G. (1968). A program for the solution of a class of geometric-analogy intelligence-test questions. In M. Minsky (Ed.), *Semantic information processing.* Cambridge: MIT Press.

Eysenck, H. J. (1998). *Intelligence: A new look*: Transaction Publishers.

Formann, A. K., & Piswanger, K. (1979). *Wiener Matrizen-Test (WMT)*: Beltz Test.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2014). Shortened Complex Span Tasks Can Reliably Measure Working Memory Capacity, *43*(2), 226–236.

Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *29*(27), 8726–8733. https://doi.org/10.1523/JNEUROSCI.2145-09.2009

Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture, *22*(3), 361–368.

Fukuda, K., Vogel, E. K., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679. https://doi.org/10.3758/17.5.673

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence Predicts Health and Longevity, but Why? *Current Directions in Psychological Science*, *13*(1), 1–4. https://doi.org/10.1111/j.0963-7214.2004.01301001.x

Green, K. E., & Kluever, R. C. (1992). Components of item difficulty of Raven's matrices. *The Journal of General Psychology*, *119*(2), 189–199.

Harrison, T. L., Shipstead, Z., & Engle, R. W. (2015). Why is working memory capacity related to matrix reasoning tasks? *Memory & Cognition*, *43*(3), 389–396. https://doi.org/10.3758/s13421-014-0473-3

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A Novel Method for Analyzing Sequential Eye Movements Reveals Strategic Influence on Raven's Advanced Progressive Matrices, *11*(10), 10.

Hilbert, S., & Stadler, M. (2017). Structural Equation Models. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–9). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1285-1

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, *10*(3), 302–323.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, *75*(3), 242–259.

Horn, J. L., & Cattell, R. B. (1966a). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253.

Horn, J. L., & Cattell, R. B. (1966b). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270. https://doi.org/10.1037/h0023816

Hossiep, R., & Hasella, M. (2010). *Bochumer Matrizentest Standard: BOMAT-Standard*: Hogrefe, Verlag für Psychologie.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(3), 340–362. https://doi.org/10.1016/0001-8791(86)90013-8

Jarosz, A. F., & Wiley, J. (2012). Why does working memory capacity predict RAPM performance? A possible role of distraction. *Intelligence*, *40*(5), 427–438. https://doi.org/10.1016/j.intell.2012.06.001

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology. General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kreuzpointner, L. (2013). Konstruktionsrational der Kurzversion des LPS-2 (LPS-2K). Retrieved from urn:nbn:de:bvb:355-epub-356934

Kreuzpointner, L., Lukesch, H., & Horn, W. (2013). *Leistungsprüfsystem 2. LPS-2. Manual*: Hogrefe.

Lee, E.-Y., Cowan, N., Vogel, E. K., Rolan, T., Valle-Inclan, F., & Hackley, S. A. (2010). Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information. *Brain: a Journal of Neurology*, *133*(9), 2677–2689. https://doi.org/10.1093/brain/awq197

Liesefeld, A. M., Liesefeld, H. R., & Zimmer, H. D. (2014). Intercommunication between prefrontal and posterior brain regions for protecting visual working memory from distractor interference. *Psychological Sscience*, *25*(2), 325–333. https://doi.org/10.1177/0956797613501170

Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology*, *5*, 1–13. https://doi.org/10.3389/fpsyg.2014.00239

Loesche, P., Wiley, J., & Hasselhorn, M. (2015). How knowing the rules affects solving the Raven Advanced Progressive Matrices Test. *Intelligence*, *48*, 58–75. https://doi.org/10.1016/j.intell.2014.10.004

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Machizawa, M. G., & Driver, J. (2011). Principal component analysis of behavioural individual differences suggests that particular aspects of visual working memory may relate to specific aspects of attention. *Neuropsychologia*, *49*(6), 1518–1526. https://doi.org/10.1016/j.neuropsychologia.2010.11.032

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*(2), 107–127. https://doi.org/10.1016/0160-2896(83)90023-5

Martínez, K., & Colom, R. (2009). Working memory capacity and processing efficiency predict fluid but not crystallized and spatial intelligence: Evidence supporting the neural noise hypothesis. *Personality and Individual Differences*, *46*(3), 281–286. https://doi.org/10.1016/j.paid.2008.10.012

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Meo, M., Roberts, M. J., & Marucci, F. S. (2007). Element salience as a predictor of item difficulty for Raven's Progressive Matrices. *Intelligence*, *35*(4), 359–368. https://doi.org/10.1016/j.intell.2006.10.001

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, *12*(2), 252–284.

Muthén, L. K., & Muthén, B. O. (2006). Mplus Version 7 user's guide. *Los Angeles, CA: Muthén & Muthén*.

Neisser, U. E., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J.,. . . others. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411–421. https://doi.org/10.1037//0278-7393.28.3.411

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working memory and intelligence--their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 61-5; author reply 72-5. https://doi.org/10.1037/0033-2909.131.1.61

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. *Variation in Working Memory*, 49–75.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378. https://doi.org/10.3758/BF03210419

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290.

Preckel, F. (2003). *Diagnostik intellektueller Hochbegabung: Testentwicklung zur Erfassung der fluiden Intelligenz* (1. Aufl.). Göttingen: Hogrefe-Verlag.

Primi, R. (2002). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence*, *30*(1), 41–70. https://doi.org/10.1016/S0160-2896(01)00067-8

Ragni, M., & Neubert, S. (2014). Analyzing Raven's Intelligence Test: Cognitive Model, Demand, and Complexity. In H. Prade & G. Richard (Eds.), *Studies in Computational Intelligence. Computational Approaches to Analogical Reasoning: Current Trends* (Vol. 548, pp. 351–370). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-54516-0_14

Raven, J. C. (1938). Progressive matrices: A perceptual test of intelligence. *London: HK Lewis and Co*.

Raven, J. C. (1940). Progressive Matrices: Instructions, Key, and Norms. *London: HK Lewis and Co*.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Rost, D. H. (2009). *Intelligenz: Fakten und Mythen*: Beltz.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning.

Schermelleh-Engel, K., & Werner, C. (2008). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 113–133). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71635-8_6

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116–141. https://doi.org/10.1016/j.jml.2014.01.004

Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory (Hove, England)*, *20*(6), 608–628. https://doi.org/10.1080/09658211.2012.691519

Spearman, C. (1927). *The Abilities of Man: Their Nature and Measurement*. London: Macmillan and Co., Limited.

Spronk, M., Vogel, E. K., & Jonkman, L. M. (2012). Electrophysiological evidence for immature processing capacity and filtering in visuospatial working memory in adolescents. *PloS one*, *7*(8), e42262. https://doi.org/10.1371/journal.pone.0042262

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. *The experimental psychology series*. Hillsdale, NJ: Erlbaum [u.a.].

Sternberg, S. (1969). Memory-scanning: mental processes revealed by reaction-time experiments. *American Scientist*, *57*(4), 421–457.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology. General*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Thurstone, L. L. (1938). *Primary Mental Abilities*. *Psychometric Monograph No. 1*: University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, *33*(1), 67–81. https://doi.org/10.1016/j.intell.2004.08.003

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. https://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392–406. https://doi.org/10.1016/j.jml.2010.02.001

Verguts, T., & DeBoeck, P. (2002). On the correlation between working memory capacity and performance on intelligence tests. *Learning and Individual Differences*, *13*, 37–55.

Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence*, *34*(3), 261–272. https://doi.org/10.1016/j.intell.2005.11.003

Vodegel Matzen, L. B.L., van der Molen, M. W., & Dudink, A. C.M. (1994). Error analysis of raven test performance. *Personality and Individual Differences*, *16*(3), 433–445. https://doi.org/10.1016/0191-8869(94)90070-1

Vogel, E. K., & Awh, E. (2008). How to Exploit Diversity for Scientific Gain: Using Individual Differences to Constrain Cognitive Theory. *Current Directions in Psychological Science*, *17*(2), 171–176. https://doi.org/10.1111/j.1467-8721.2008.00569.x

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/nature04171

Wechsler, D. (2008). Wechsler adult intelligence scale-Fourth Edition (WAIS-IV). *San Antonio, TX: NCS Pearson*, *22*, 498.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. H. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 256–263. https://doi.org/10.1037/a0021613

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433. https://doi.org/10.3389/fpsyg.2013.00433

Wongupparaj, P., Kumari, V., & Morris, R. G. (2015). The relation between a multicomponent working memory and intelligence: The roles of central executive and short-term storage functions. *Intelligence*, *53*, 166–180. https://doi.org/10.1016/j.intell.2015.10.007

Zimmer, H. D. (2008). Visual and spatial working memory: From boxes to networks. *Neuroscience and Biobehavioral Reviews*, *32*(8), 1373–1395. https://doi.org/10.1016/j.neubiorev.2008.05.016

# Lebenslauf

## Persönliche Informationen

| | |
|---|---|
| Nachname | Krieger (geb. Domnick) |
| Vorname | Florian |
| Geburtsdatum | 23. März 1991 |
| Geburtsort | Berlin |
| Adresse | Lilienstraße 22, 66119 Saarbrücken |
| Telefon | 01523 3531686 |
| E-Mail | f.domnick@mx.uni-saarland.de |
| Familienstatus | verheiratet |
| Kinder | Antonia Krieger (*07.04.2017) |

## Abitur und universitäre Ausbildung

| | |
|---|---|
| 2009 | Abitur<br>*Landesgymnasium für Hochbegabte, Schwäbisch Gmünd* |
| 2009 – 2012 | Studium der Psychologie, Bachelor of Science<br>*Universität des Saarlandes, Saarbrücken* |
| 07/2012 | Abschluss Bachelor of Science |
| 2012 – 2014 | Studium der Psychologie, Master of Science,<br>*Universität des Saarlandes, Saarbrücken* |
| 04/2013 – 03/2014 | Fast-Track-Programm für den M.Sc. als Qualifikation für die Promotion<br>*IRTG Adaptive Minds, Universität des Saarlandes, Saarbrücken* |
| 03/2014 | Abschluss Master of Science |
| Seit 04/2014 | Promotion am internationalen Graduiertenkolleg “Adaptive Minds”<br>*IRTG GRK 1457, Universität des Saarlandes, Saarbrücken* |

## Berufliche Tätigkeiten

| | |
|---|---|
| 04/2014 – 03/2017 | Wissenschaftlicher Mitarbeiter<br>*Universität des Saarlandes, Saarbrücken* |
| Seit 05/2017 | Wissenschaftlicher Mitarbeiter<br>*Université du Luxembourg, Esch-zur-Alzette, Luxembourg* |