Why Pictures are Beneficial to Learning: Contrasting Explanations of the Multimedia Effect

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Erlijn Wendela van Genuchten, M.Sc.
aus Geldrop (NL)

Tübingen
2013

Tag der mündlichen Qualifikation:	18.10.2013
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Katharina Scheiter
2. Berichterstatter:	Prof. Dr. Stephan Schwan

Acknowledgements

I would like to thank…

- Katharina Scheiter and Anne Schüler for being my supervisors and providing helpful suggestions for improvement,
- Stephan Schwan for being my second reviewer,
- my research assistants Sarah Paganini, Raphael Frank, Gabriel Kornwachs, and Marion Kornmayer for helping me with data gathering and scoring,
- my parents Jos and Sabine, my brother Maarten, and Marina for supporting me through good and through bad times,
- Peter Cheng and Paul Kirschner, who never stopped believing in me,
- my Dutch friends Roos Verboog and Iris Visser, and my German friends Hannah Greving, Jana Prokoph, Moni Lenk, and Brigitte Schreier for making me feel good every time I met them,
- Charlotte van Hooijdonk for our pleasant and productive collaboration,
- my former colleagues at the Knowledge Media Research Center (Institut für Wissensmedien)
    - Birgit Imhof, Jana Arndt, Carina Schubert, Mareike Wollenschläger, Jessica Kornmann, Anja Rudat, Laure Philippon, Yvonne Kammerer, and Halszka Jarodzka for having many great chats at the institute and beyond,
    - Alexander Eitel for being my faithful roommate,
    - Torsten Kurbad and Sebastian Groteloh for supporting me many times with technical issues,
- and my current colleagues at the Stuttgart Media University (Hochschule der Medien), Mathias Hinkelmann, Tobias Seidl, Sebastian Kelle, Andrea Hempel, and Katrin Sauermann, for supporting me to get through the last stages of this dissertation.

Table of contents

1 INTRODUCTION 7

1.1 The Multimedia Effect 8

1.2 Explaining the Multimedia Effect: Cognitive Theory of Multimedia Learning 11

1.3 Explaining the Multimedia Effect: Account of Computational Offloading 13

1.3.1 Information processing when learning with text only 13

1.3.2 Information processing when learning with text and pictures 14

1.4 Differences between Cognitive Theory of Multimedia Learning and Account of Computational Offloading 15

1.5 Task Types 17

1.5.1 Conceptual tasks 18

1.5.2 Causal tasks 19

1.5.3 Procedural tasks 21

1.5.4 Linking theoretical accounts to task type 24

1.6 Working Memory Involvement during Multimedia Learning 26

1.6.1 Structure of working memory 26

1.6.2 Assessing working memory involvement 27

1.6.3 The role of working memory during multimedia learning 29

1.6.4 Discrepancy between computational offloading account, recent working memory research and empirical results 32

1.6.5 Linking theoretical accounts to working memory 33

1.7 Hypotheses 34

2 EXPERIMENT 1 36

2.1 Method 36

2.1.1 Participants and design 36

2.1.2 Materials and measures 37

2.1.3 Procedure 39

2.1.4 Data analysis 40
2.2 Results 41
2.2.1 Free recall 41
2.2.2 Recall verification accuracy 43
2.2.3 Transfer verification accuracy 45
2.2.4 Integration verification accuracy 49
2.3 Discussion 50

3 EXPERIMENT 2 53

3.1 Method 53
3.1.1 Participants and design 53
3.1.2 Materials and measures 54
3.1.3 Procedure 55
3.1.4 Data analysis 55
3.2 Results 56
3.2.1 Free recall 56
3.2.2 Recall verification accuracy 57
3.2.3 Free transfer 58
3.2.4 Transfer verification accuracy 58
3.2.5 Integration verification accuracy 59
3.2.6 Study times 62
3.3 Discussion 62

4 EXPERIMENT 3 65

4.1 Method 65
4.1.1 Participants and design 65
4.1.2 Materials and measures 66
4.1.3 Procedure 67
4.1.4 Data analysis 68
4.2 Results 69
4.2.1 Main effects 69
4.2.2 Two-way interactions 69

4.2.3 THREE-WAY INTERACTIONS 71
4.3 DISCUSSION **73**

5 GENERAL DISCUSSION **76**

5.1 SUMMARY OF THE RESULTS AND INTERPRETATION **77**
5.2 FACTORS INFLUENCING THE PATTERN OF RESULTS **78**
5.2.1 FACTORS AFFECTING IMAGERY DURING LEARNING 78
5.2.2 FACTORS AFFECTING THE ASSESSMENT OF WORKING MEMORY INVOLVEMENT DURING LEARNING 80
5.2.3 FACTORS AFFECTING THE EXPRESSIVENESS OF THE LEARNING OUTCOME MEASURES 82
5.3 LIMITATIONS **85**
5.4 STRENGTHS **87**
5.5 IMPLICATIONS **88**
5.5.1 THEORETICAL IMPLICATIONS 89
5.5.2 PRACTICAL IMPLICATIONS 90
5.6 CONCLUDING REMARKS **91**

6 SUMMARY **92**

7 ZUSAMMENFASSUNG **93**

8 REFERENCES **94**

9 LIST OF FIGURES **102**

10 LIST OF TABLES **104**

1 Introduction

In the last few decades, the use of multimedia has become ubiquitous due to the increased deployment of computers and the internet. Multimedia is the combination of the different presentation formats such as written or spoken text and pictures, animations, or videos (Mayer, 2005a). These days, not only for example school books but also websites and digital learning environments allow the presentation of different presentation formats. The scientific field of multimedia learning focuses on investigating how learning material in different presentation formats (i.e., combinations of different external representations) can support deeper information processing, understanding of information, and knowledge construction (i.e., construction of internal/mental representations; Mayer, 2009a). An important effect in the field of multimedia learning is the multimedia effect. According to the multimedia effect, learning is fostered by adding pictures to text (Mayer, 2009b). A lot of research has confirmed this effect (see Section 1.1); however it is still unclear *why* pictures are beneficial to learning. Therefore, in this dissertation, two accounts, the most common account of Cognitive Theory of Multimedia Learning (Mayer, 2005b) and the alternative account of computational offloading (Larkin & Simon, 1987), are contrasted to see which account explains the multimedia effect best. After describing the multimedia effect in more detail (see Section 1.1), these theoretical accounts are described (see Section 1.2 and Section 1.3). These descriptions include explanations of why these theoretical accounts assume that pictures are beneficial to learning. Also, theoretical implications derived from these accounts are explained, which pertain to the role of the importance of visuo-spatial information for understanding the learning material and the involvement of working memory during learning (see Section 1.5.4 and 1.6.5). By testing these implications, insight into which theoretical account explains the beneficial effect of pictures best and therefore into why pictures are beneficial to learning can be obtained. These insights not only contribute to learning theories that describe and predict behaviour during multimedia learning (cf. Mayer, 2009a), but also potentially to instructional practice by being able to more carefully design multimedia learning material.

1.1 The Multimedia Effect

The multimedia effect is a well established effect, as a broad range of studies using different tasks and learning outcomes has confirmed this effect (for reviews see Anglin, Vaez, & Cunningham, 2004; Carney & Levin, 2002; Mayer, 2009b). An example of the multimedia effect is that it is easier to understand how a bicycle pump works when a picture depicting the inner components and the states of these components during pumping accompanies the text than when the functioning of the bicycle pump is described by text only (e.g., Mayer, 2005a; 2009b). The typical design of studies that test the multimedia effect involves two conditions in which participants learn either with verbal information only (i.e., control condition) or with verbal and pictorial information (i.e., experimental condition). The knowledge participants obtain from the learning material is typically assessed immediately after learning. Within this design, it is possible to vary certain aspects, such as the type of learning task and the type of learning outcome measures. For example, studies investigating the multimedia effect have used a broad range of tasks with different types of learning contents, such as the classification of sail boats (Wilcox, Merrill, & Black, 1981), the functioning of a bicycle pump (Mayer & Gallini, 1990), or how to bandage a hand (Michas & Berry, 2000). Also, learning outcomes have been measured using different performance measures, such as comprehension (Hannus & Hyönä, 1999), creative problem solving (Mayer & Gallini, 1990), or bandaging performance (Michas & Berry, 2000).

In Table 1, experiments that tested the multimedia effect, using different types of tasks and different types of learning outcomes, are listed. Studies were selected based on their learning materials, which are similar to the learning materials used in this dissertation. For all experiments, participants learning with a written text only were compared to participants learning with a written text and a static picture. The effect size Cohen’s d was calculated for each learning outcome measure, which ranged from -0.05 to 1.93 — with one exception of -1.58 — with a mean effect size of 0.66, which is a medium to large effect (Cohen, 1988).

Table 1

Effect Sizes for the Multimedia Effect in Previous Research

Reference	Task	Learning outcome measures (Original variable name: Cohen's d*)
Hannus & Hyönä, 1999, exp. 2	Description of animals	Recall: - Main points: 0.10 - Detailed points: 0.33 Transfer: - Comprehension: 0.23
Wilcox, Merrill, & Black, 1981	Classification of sail boats	Transfer: - Classifying unencountered instances of concepts: 0.54
Mayer & Gallini, 1990, exp. 1	The functioning of brakes	Recall: - Explanative information: 1.93 - Non-explanative information and verbatim recall: NS Transfer: - Creative problem solving: 1.78
Mayer & Gallini, 1990, exp. 2	The functioning of pumps	Recall: - Explanative information: 1.80 - Non-explanative information: -1.58 - Verbatim recall: NS Transfer: - Creative problem solving: 1.33
Mayer & Gallini, 1990, exp. 3	The functioning of generators	Recall: - Explanative information: 1.30 - Non-explanative information: NS - Verbatim recall: 0.51

		Transfer: - Creative problem solving: 1.77
Moreno & Valdez, 2005, exp. 1	The formation of lightning	Recall: - Recall: 0.40 Transfer: - Transfer: 0.52
Nam & Pujari, 2005	Refrigeration cycle	Transfer: - Transfer: 0.29
Waddill & McDaniel, 1992, exp. 1	Crumbling cliffs	Recall: - Target Detail: -0.05 - Target Relational: 0.38
Michas & Berry, 2000, exp. 1	Bandaging a hand	Recall: - Bandaging performance: 1.15 - Questions: 0.86
Michas & Berry, 2000, exp. 3	Bandaging a hand	Recall: - Bandaging performance: 1.64 - Questions: 0.18
Stone & Glock, 1981	Assembly of a model loading cart	Recall: - Assembly errors 1.14**

* A positive d-value implies a multimedia effect.

** As this measure involves errors, the sign of the effect size has been inversed.

Although the multimedia effect is a well established effect, and therefore pictures are said to be beneficial to learning, additional clarity is still sought to describe *why* pictures are beneficial to learning. Several theoretical accounts have been proposed to explain why pictures are beneficial to learning. The Cognitive Theory of Multimedia Learning (Mayer, 2005b) states that pictures are beneficial to learning, because not only a verbal but also a pictorial mental representation — and ideally an integrated mental representation — are yielded in long-term memory. This account is different from the account of computational offloading (Larkin & Simon, 1987), which posits that pictures facilitate information processing during learning, especially when visuo-spatial information is important for understanding the learning task.

1.2 Explaining the Multimedia Effect: Cognitive Theory of Multimedia Learning

The most prominent theoretical account in the field of multimedia learning, the Cognitive Theory of Multimedia Learning (CTML; see Figure 1), explains the multimedia effect by positing that learning from text and pictures yields a richer mental representation than learning from text only (Mayer, 2005b). CTML postulates that during learning with text and pictures, relevant information is selected from verbal and pictorial information that enters sensory memory (e.g., through the eyes). The selected information is further processed in working memory, which is responsible for processing verbal and pictorial information and is assumed to have a limited capacity to process information at the same time (Baddeley, 1986). According to Baddeley (2007), working memory is a temporary storage system in memory that enables complex thought. When the selected information is verbal (e.g., from listening to spoken text), this information is processed in the word sound base[1]. According to CTML, the word sound base corresponds to Baddeley's working memory subsystem that is responsible for processing verbal information, namely the phonological loop. Words that are processed in the word sound base are organised in a verbal mental representation. Analogously, when the selected information is pictorial (e.g., from looking at static pictures), this information is processed in the visual image base. According to CTML, the visual image base corresponds to Baddeley's working memory subsystem that is responsible for processing visuo-spatial information, namely the visuo-spatial sketchpad. Images that are processed in the visual image base are organised in a pictorial mental representation. It is important to note that CTML does not take contemporary research into account that shows that text containing visuo-spatial information is also processed in the visuo-spatial sketchpad (see Section 1.6.1; e.g., Gyselinck, De Beni, Pazzaglia, Meneghetti, & Mondoloni, 2007). The verbal and pictorial mental representations are integrated with each other and with prior knowledge that is retrieved from long-term memory. This integrated mental representation is more than the sum of the purely verbal and purely pictorial representation, because it is thought to reflect a full understanding of the learning material. Accordingly, it does not only allow for recall of

[1] CTML posits that spoken text is processed in the word sound bases, whereas written text is processed in the visual image base. This assumption has been criticized as it is not in line with Baddeley's (1986) working memory model (Rummer, Schweppe, Scheiter, & Gerjets, 2008).

information (i.e., remembering information) but also for transfer of knowledge to novel situations (i.e., reasoning about the learning material; Mayer, 2005a).

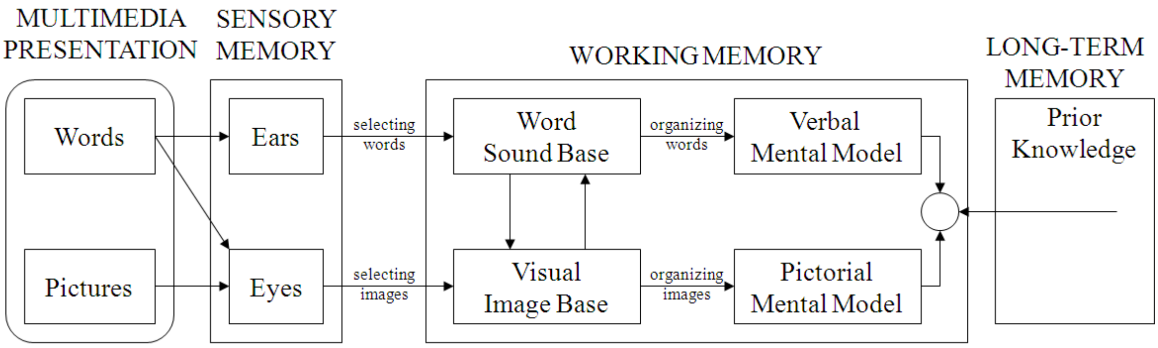

Figure 1. Cognitive Theory of Multimedia Learning (obtained from Mayer, 2005b).

CTML is grounded in Paivio's (1991) dual coding theory, according to which processing information from text and pictures yields two mental representations, namely a verbal and a non-verbal representation. According to dual coding theory, text elements are stored as logogens in the verbal representation, whereas picture elements are stored as imagens in the non-verbal representation. When learning with text and pictures, the chance that two — instead of one — representations of the same information are constructed is larger than when learning with text only. Accordingly, the quantity of information stored in long-term memory increases when learning with text and pictures. As information can not only be directly activated by sensory input, but also by spreading activation between and within these mental representations, having two representations in memory increases the likelihood that the correct logogen or imagen is activated and therefore increases the chance that information is recalled.

In reference to dual coding theory, CTML predicts that pictures are beneficial to learning, because studying both text and pictures increases the chance that two representations are constructed in memory than when learning with text only, based on which an integrated representation can be constructed. CTML does not mention differences in dual coding between different types of information; thus it can be assumed that this beneficial effect is predicted irrespective of the learning content.

In contrast to this explanation that relies on the construction of a richer internal representation and facilitated information retrieval, an alternative explanation of the multimedia effect focuses on facilitated information processing during learning, especially when processing visuo-spatial information. This alternative is addressed by the account of computational offloading (Larkin & Simon, 1987).

1.3 Explaining the Multimedia Effect: Account of Computational Offloading

The alternative account of computational offloading proposes that pictures are beneficial to learning, because they require fewer cognitive resources for information processing during learning, and hence mental model construction, than text (Larkin & Simon, 1987). This is especially the case when visuo-spatial information is processed (Ainsworth, 1999; Larkin & Simon, 1987; Scaife & Rogers, 1996). According to contemporary theories of learning and memory, information processing during learning takes place in working memory. Thus, it can be argued that the computational offloading approach suggests that pictures enable more efficient use of working memory resources. Considering the focus of the computational offloading approach on the processing of visuo-spatial information, it can be argued that offloading should be observable especially for the working memory system that is responsible for processing visuo-spatial information, namely the visuo-spatial sketchpad (which will be introduced in Section 1.6.1). In the following, assumptions derived from the account of computational offloading concerning how information is processed and knowledge is constructed when learning from text only versus learning with text and pictures are contrasted.

1.3.1 Information processing when learning with text only

Information processing when learning with text only initially involves reading words and sentences (e.g., reading "A plate is put on the table, a knife and fork are put next to it."). These words and sentences have to be interpreted (e.g., fork and knife are cutlery needed to eat the food on the plate; Scaife & Rogers, 1996) and visual search is applied to find and group text elements that belong together (e.g., the fork, knife, and plate belong together; cf. Larkin & Simon, 1987). After elaborating on the meaning of the text and grouping elements, these words and sentences can be used to elaborate on the representation that the text concerns (e.g., the fork, knife, and plate on the table show that the breakfast table is set; Denis, 1989; Kintsch & Van Dijk, 1978). Without any reliance on external visual support, the understanding of this representation is used to create a mental representation (e.g., mental representation of a breakfast table that is set; Pearson, de Beni, & Cornoldi, 2001). This mental representation is stored in long-term memory.

To be able to add meaning to the text and create a mental representation, imagery is said to play an invaluable role (de Vega, Cocude, Denis, Rodrigo, & Zimmer, 2001; ; Sadoski, 2001; Sadoski & Paivio, 2004). By means of imagery, meaning is added to text — especially with regard to visuo-spatial information that reflects visuo-spatial information in the real world — through generating or recalling mental images, which reflect visuo-spatial information from the real world, from long-term memory (e.g., image of a knife; Glasgow & Papadias, 1995), through modifying these mental images (e.g., modify the image of the knife so that it reflects a knife that is used for breakfast), and through applying operations such as image generation, maintenance, inspection, manipulation, etc. to these images (e.g., mirroring the image of the knife; Kosslyn, 1999). According to Glenberg, Kruley, and Langston (1994), these mental images can result in a mental representation (e.g., a mental image of the breakfast table).

Although imagery allows creating a mental representation based on the text, it requires a lot of cognitive resources to apply visual search, to interpret text (Larkin & Simon, 1987), to recall mental images, and to apply operations to these images to create a mental representation (Kosslyn, 1999). For example, Kirsh and Maglio (1994) found that participants solved Tetris problems more quickly, easily, and reliably when looking at Tetris-pieces that they had turned on the screen than when imagining Tetris-pieces that they had turned in their heads. As processing pictures does not require recalling mental images and applying operations to these images, picture processing is likely to be less error prone and reduces the amount of cognitive effort required to understand the learning material.

1.3.2 Information processing when learning with text and pictures

Information processing when learning with text and pictures, initially involves automatically perceiving visuo-spatial information from pictures (e.g., perceiving a round-shaped object; Larkin & Simon, 1987). This relevant visuo-spatial information requires the recognition of the depiction as opposed to the interpretation of words and sentences (e.g., the object is recognised as a plate; Larkin & Simon, 1987). Visual search is required only to a small extent as related elements can be grouped in pictures (e.g., the fork is depicted next to the plate; Larkin & Simon, 1987). After recognising the visuo-spatial information, a mental representation can be created, whereby the picture can be seen as an analogical — that is, structure-preserving — expression of a mental representation (e.g., mental representation of a breakfast table that is set; Glenberg & Langston, 1992, Gyselinck & Tardieu, 1999).

When extracting visuo-spatial information from a picture and creating a mental representation, visuo-spatial perception, instead of imagery, allows directly perceiving, recognising, and using information from the picture. This information is used to add meaning to the text (e.g., seeing at a breakfast table that is set; de Vega et al., 2001; Kosslyn, 1999; Scaife & Rogers, 1996; Zhang & Norman, 1994).

Compared to imagery, the process of visuo-spatial perception is less error prone and requires fewer cognitive resources than mental imagery, as processing pictures does not require recalling mental images and applying operations to these images due to external visual support (cf. Glasgow & Papadias, 1995). Therefore, pictures are assumed to be a transitory step in transforming information into mental representations (Glenberg et al., 1994) and are said to be computational offloading (Larkin & Simon, 1987). Pictures thus facilitate the construction of a mental representation (Gyselinck & Tardieu, 1999) and consequently elicit a multimedia effect.

1.4 Differences between Cognitive Theory of Multimedia Learning and Account of Computational Offloading

From these descriptions, it can be summarised that CTML explains the multimedia effect based on a richer mental representation in long-term memory, whereas the account of computational offloading explains the multimedia effect based on facilitated information processing during learning. From these foci, three importance distinctions can be derived.

One important distinction between both theoretical accounts is that only the account of computational offloading predicts that the beneficial effect of pictures depends on the amount of visuo-spatial information that needs to be processed, that is, how important visuo-spatial information is for understanding the learning material. As visuo-spatial information can be perceived more easily from pictures than from text, it can be argued that pictures are especially computational offloading when visuo-spatial information is conveyed (cf. Larkin & Simon, 1987; Scaife & Rogers, 1996). This would imply that pictures are more computational offloading the more important information visuo-spatial information is for understanding the learning task. In contrast, CTML does not make any predictions concerning differences in mental representations based on the importance of visuo-spatial information. Accordingly, the difference between CTML and the account of computational offloading concerning the importance of visuo-spatial information can be investigated by comparing the multimedia effect in task types that differ in how important visuo-spatial information is for understanding the learning material. The account of computational offloading would predict a larger multimedia effect the more important visuo-spatial information is for understanding, whereas CTML would not predict differences in the size of the multimedia effect between task types.

A second important distinction between both theoretical accounts is that the account of computational offloading predicts that differences in information processing during learning account for the differences between task types. As according to Baddeley (1986) working memory is responsible for information processing during learning and task types can be distinguished based on how important visuo-spatial information is for understanding the task, it can be argued that cognitive resources in working memory are used to a different extent depending on the type of task. In contrast, although working memory also plays an important role in CTML, CTML does not make predictions concerning information processing in working memory depending on the type of task. Accordingly, this difference between CTML and the account of computational offloading can be investigated by comparing working memory involvement in different types of tasks when learning with text and pictures. The account of computational offloading would predict differences in information processing between task types, whereas CTML would not predict such differences.

A third important distinction between both theoretical accounts is how working memory is assumed to be involved when learning with text only versus text and pictures. The account of computational offloading argues that pictures are computational offloading, implying that processing text only requires more cognitive resources than processing pictures. In contrast, CTML predicts that picture processing requires more visuo-spatial information processing than text processing, as CTML does not take contemporary research concerning the role of the visuo-spatial sketchpad during text processing into account (see also Section 1.6.1). Accordingly, this difference between CTML and the account of computational offloading can be investigated by comparing working involvement when learning with text only versus learning with text and pictures. The account of computational offloading would predict that working memory involvement is higher when learning with text only than when learning with text and pictures, whereas CTML would predict the opposite.

To summarise, the three important differences between CTML and the account of computational offloading pertain to their predictions concerning 1) whether the beneficial effect of pictures depends on the amount of visuo-spatial information processing that is required by a certain task type, 2) whether information processing in working memory differs between task types, and 3) whether information processing in working memory differs between processing text only or processing text and pictures.

In the following, the concept of task type, which concerns the use of visuo-spatial information in different types of tasks, and the concept of working memory, which concerns information processing during learning, are introduced in more detail. These concepts are used to investigate the differences between CTML and the account of computational offloading and contrast these accounts.

1.5 Task Types

In this dissertation, the concept of task type is used to distinguish between tasks that differ in how important visuo-spatial information is to understand the learning material. In multimedia research based on CTML, evidence for the beneficial effect of pictures has often been treated equally, even though tasks convey different types of information and therefore differ in how important visuo-spatial information is for understanding the learning task. For example, pictures have been used to illustrate how domains are organised in terms of their constituting elements and interrelations, how mechanical systems work, or how to perform a certain procedure. In line with Reigeluth and Stein (1983), I will refer to tasks conveying these types of information as conceptual, causal, and procedural tasks, respectively. It is yet unclear whether the multimedia effect is equally likely to appear for these task types, as could be expected based on CTML. Certainly, there are multimedia studies for each task type; however, these studies differ on a variety of dimensions (e.g., type of picture used, difficulty of materials, learning outcome measures; see Table 1) making it difficult to make a definite statement concerning the size of the multimedia effect for these task types. This issue can be addressed by studying the multimedia effect for the different task types within one experiment. In the following, for each task type, defining features are identified, an example taken from published multimedia research is described, and how information is distributed across text and pictures is discussed. These descriptions show how task types differ in their use of visuo-spatial information to represent task-relevant information.

1.5.1 Conceptual tasks

In conceptual tasks, conceptual structures are described and depicted (Hiebert & Carpenter, 1992; Reigeluth & Stein, 1983). A conceptual structure reflects (instances of) concepts and their relationships. Concepts relate to a set of objects, symbols, or events, which are composed of and can be decomposed into their defining characteristics and relationships between those characteristics (Mervis & Rosch, 1981). When learners have learned a concept, they are able to identify, recognize, classify, describe, and make predictions about instances of this concept, and are able to assign previously unknown entities to a known concept (Reigeluth & Stein, 1983). In research on the effect of combining text and pictures, conceptual tasks were for example used by Hannus and Hyönä (1999). Hannus and Hyönä used textbook passages concerning snakes, hawks, lizards, grasshoppers, ferns, and birds nesting in holes. Most pictures showed the general structure of an organism or showed two organisms that resembled each other to point out their interrelations (e.g., discriminating features).

In the conceptual tasks used in Experiment 1 and 2 (see Figure 2 for an example), the text conveyed information concerning non-observable, non-visual aspects (e.g., in the text was described that the person who is called Mr. Blue Circle is the Mayor of the town Bandelop) and also concerning the nature of relationships between objects (e.g., it was described in the text that Miss Pink Heart is said to have an affair with Mr. Purple Arc), whereas the pictures conveyed visual aspects and the presence of relationships amongst objects (e.g., the picture showed that Miss Pink Heart is related to two other people in town), however not the nature of these relationships (e.g., it was not shown in the picture *how* Miss Pink Heart is related to the two other people). Importantly, in conceptual pictures, the spatial arrangement of elements typically does not reflect spatial arrangement of elements in the real world (e.g., the physical distance in the picture between Miss Grey Cloud and Mr. Yellow Rectangle was not the same as their distance in the real world). Instead, space is used in a metaphorical way, meaning that space represents conceptual relatedness of elements on more abstract levels (e.g., the physical distance in the picture showed that Mr. Blue Circle is more closely related to Mrs. Green Rectangle than to Miss Pink Heart). Therefore, the feature to which this relatedness refers can typically not be read off from the picture, but has to be inferred from the text.

Bandelop is a village in Pandanstan. The village has 17 inhabitants. Mr Blue Circle is the Mayor and has four children. Yellow and Green Circle are siblings and are 8 years old. Mr Circle's oldest son often plays with Purple. He is the baker's son. Purple also has some other friends. Apart from that, Orange is Purple's girlfriend. She is living on the opposite side of the street. Her mum often visits the baker and helps him with cleaning the house. In the meantime, he makes cake and bread for the other people living in the village. There is a rumour saying that they are having an affair. Another rumour says that the music school teacher, Mrs Note, is interested in Mrs Triangle. However, Mrs Note has a husband, Mr Hexagon. Therefore, this rumour is probably not true. A fourth woman is living in the village, Ms Grey Cloud. She is homeless and is ignored by the others. In spite of that, she feels at ease in Bandelop. She has been chased off from other villages in the past. She has built her cardboard house next to the sports field. That's where Mr Hexagon and Mr Arch are playing headball every Sunday morning. Once a year, all inhabitants gather for a banquet, except of course Ms Grey Cloud. On December 31th they celebrate the old year and the coming of the new.

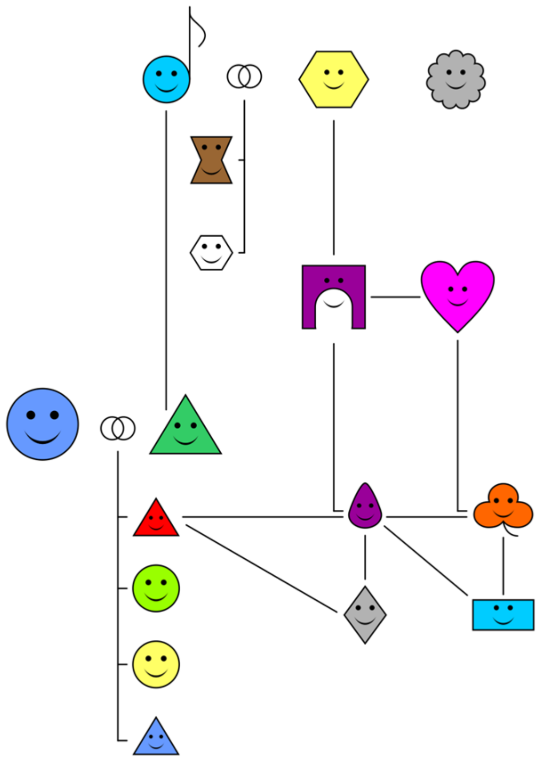

Figure 2. Example of a conceptual task used in Experiment 1 and 2 (original version was in German).

1.5.2 Causal tasks

Causal tasks are probably the most frequently used tasks in multimedia research. In causal tasks, cause-and-effect chains are described and depicted. A cause-and-effect chain reflects a change in something and the consequence on something else (Mayer, 2009b; Reigeluth & Stein, 1983). When a learner has learned a causal relationship, s/he is able to identify why something happened (i.e., cause) and what will happen as a result of this cause (i.e., effect; Reigeluth & Stein, 1983). In previous research on the effect of combining text and pictures, causal tasks were for example used by Mayer and Gallini (1990) who used learning material in which text and/or pictures conveyed the causal chain of events leading to a bike tyre being pumped.

In the causal tasks used in Experiment 1 and 2 (see Figure 3 for an example), the text and the picture conjointly conveyed information concerning the important components of the system, that is, its configuration and the underlying kinematics (Hegarty & Just, 1993). Components were typically mentioned in the text (e.g., it was described in the text that the bucket is part of some sort of balance) and their visuo-spatial appearance and arrangement was shown in the picture (e.g., the picture showed that the bucket is attached to a rope and that a stone is attached to the other end). Regarding a causal system's kinematics, text was typically used to explain the underlying principles (e.g., it was described in the text that a stone travels a long way through the air and falls into a container), whereas the spatial arrangement of components provided information concerning the functional relationships amongst the components (i.e., how the components were connected and thus how changes in one component would affect another; e.g., a twig in front of the stone stops it from rolling down the hill). Importantly, in causal pictures, the spatial arrangement of components does not represent conceptual relatedness as was the case for conceptual pictures; instead, the spatial arrangement best reflects the spatial layout as observed in the real world (i.e., physical similarity), which was also relevant to the system's functioning (e.g., the egg above the bucket is also above the bucket in real life and shows that the egg can roll into it). In many cases, and also in the present materials, arrows are used to help learners to better understand the kinematics of the system and to infer causality in causal tasks (e.g., Heiser & Tversky, 2006; Imhof, Scheiter, & Gerjets, 2011; Münzer, Seufert, & Brünken, 2009; Tversky, Heiser, Lozano, MacKenzie, & Morrison, 2008).

One of the things people in Pandanstan eat is eggs. The birds who lay the eggs shouldn't be disturbed and therefore live a long way from the village. Because of that, the inhabitants of a village in Pandanstan invented a machine to give a notification when there are enough eggs to be collected. When the birds lay their eggs in their nests, the eggs roll into a grey bucket. This bucket is part of some kind of balance. With an increasing number of eggs in it, the bucket sinks towards the earth. When the bucket hits the ground, it makes a wheel turn around very slowly. However, this movement is enough to make a twig move from its original position. This twig stops a large stone from rolling down a hill. So by removing the twig, it is possible for the stone to roll down the hill. This stone hits a larger stone. This makes a small stone be catapulted into the air. Because of the size of the large stone, this smaller stone travels a long way through the air. This is necessary because the birds live far away from the village. In the village, the stone falls into a container. This container has several compartments. Depending on the compartment in which the stone falls, it is then decided which farmer in the village has to collect the eggs.

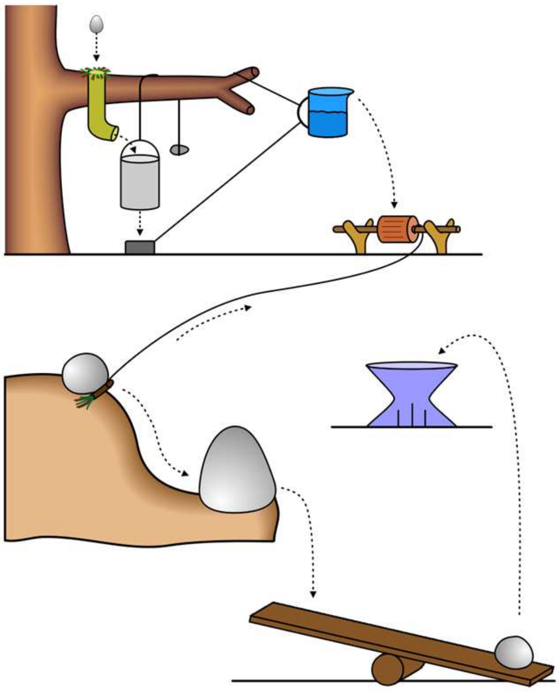

Figure 3. Example of a causal task used in Experiment 1 and 2 (original version was in German).

1.5.3 Procedural tasks

In procedural tasks, procedural structures are depicted and described. A procedural structure reflects the temporal order and spatial relationships between actions that are intended to achieve a goal (Arguel & Jamet, 2009; Brunyé, Taylor, Rapp, & Spiro, 2006; Reigeluth & Stein, 1983). A procedure is often referred to as a skill, a technique, or a method. When learners have learned a procedure, they know how to do something (Reigeluth & Stein, 1983). Some procedural tasks convey procedures by depicting only the transformations of elements (i.e., showing a sequence of states as a result of the performed actions). In previous research on the effect of combining text and pictures, a procedural task conveying *states* was for example used by Stone and Glock (1981). They used instructional material in which a text and/or picture conveyed the states when assembling a cupboard. Other procedural tasks convey *actions* by describing/depicting the interactions between body parts and elements. A procedural task conveying actions was for example used by Michas and Berry (2000). They used instructional material in which a text and/or picture described/depicted the actions required for bandaging a hand. In the studies reported in this dissertation, both types of procedural tasks were used.

In the procedural tasks used in Experiment 1 and 2 (see Figure 4 for an example) in which states were depicted, multimedia materials resembled that of causal tasks in several ways. Specifically, the text and the picture conjointly conveyed information about the important elements (i.e., what was subject to the action) and how these elements were transformed (e.g., the text and picture conveyed a tree that becomes bent). These elements were typically mentioned in the text (e.g., it was described in the text that this tree is called a Beetree), and their visuo-spatial appearance and arrangement was shown in the picture (e.g., the picture showed that the Beetree is brown and has a triangle shaped trunk), as with components in a causal task. Importantly, in procedural pictures, elements are often shown multiple times to illustrate the changes in the elements (e.g., in the picture the Beetree was shown several times to illustrate changes in its form) due to the actions that are performed (e.g., Brunyé et al., 2006; Stone & Glock, 1981). Regarding the transformations of these objects, text is typically used to explain the underlying principles and goals (e.g., it was explained in the text that the Beetree needs to be bent to obtain drinkable juice), whereas the spatial arrangement of objects and their varying states provides information concerning causal and temporal relationships (e.g., the picture showed *how* the tree needs to be bent so that the juice can be obtained). Moreover, the spatial layout of multiple pictures is used to reflect the temporal order of actions. Here, earlier element states precede subsequent states.

Beetree juice is a delicacy in Pandanstan. It can be obtained from a small tree called Beetree. However, a careful bending of the tree is necessary to get drinkable juice. First, the trunk cables are fixed in the ground, in so that the force to bend has an optimal effect. When they are both fixed, they are brought up to the top of the tree, in equal sized spirals and parallel. Then the tree is bent so that it has the shape of a wave. It is important that the branches spring from the inner side of a bow. This is only possible when the right distance between the cables is chosen. The angle between the trunk and the cable should be 45 degrees. By using this angle, the tree trunk doesn't break. After bending the trunk, the tree has to be left untouched for one month. After that, the cables are removed. In the meantime, blisters that contain juice grow on the trunk. Barrels are put on the ground to collect the juice. Then a hole is punched in these blisters. This allows the Beetree juice to drip out of the blisters. The barrels are collected after two hours.

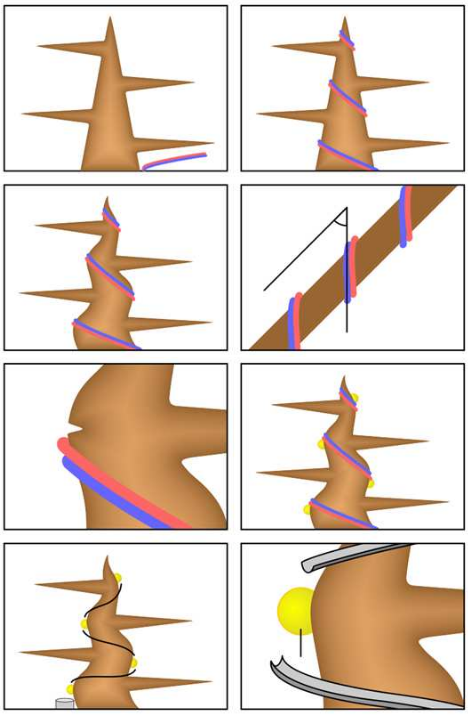

Figure 4. Example of a procedural task used in Experiment 1 and 2 (original version was in German).

In the procedural tasks used in Experiment 3 (see Figure 5 for an example) in which pictures depicted actions, visuo-spatial information was used in the same way, except that the text and pictures also conveyed body parts of the actor performing the actions (e.g., the text and picture conveyed to put your arm under the victim's armpit). As a consequence, kinaesthetic information that involves information about the position and movement of the actor's body parts required to perform the actions was included (e.g., it was conveyed that when sitting down in squad position, a foot should be placed on either side of the victim's body; cf. Ryle, 2000). Another difference was that text included action-related words to describe the actions that had to be performed, which are assumed to trigger imagery (e.g., kneel at the side of the victim; Fischer & Zwaan, 2008). Therefore, procedural tasks describing actions can be assumed to trigger imagery, whereas procedural tasks describing states do not. When imagery is triggered, this does not automatically imply that the process of imagery is successful, as imagery requires a lot of mental resources and is error prone (as is discussed in Section 1.3.1). Especially when visuo-spatial information in 3D space is conveyed, imagery could be more error prone than looking at a picture.

Rautek

Kneel at the left side of the victim. Put your right foot behind his head.

Put your right hand in his right armpit. Put your left hand from the back in his left armpit. Make the victim sit down with a continuous move. Put your body behind the victim. Put your arms under his armpits. Put one lower arm of the victim in horizontal position in front of his chest. Put your hands with closed fingers and thumbs around the lower arm. Sit down in squat position with your feet at each side of the victim and as close as possible to the victim.

Lift the victim by stretching your legs. Carry the victim backwards away from the spot.

Figure 5. Example of a procedural task used in Experiment 3 (original version was in Dutch; obtained from Henny, 2006).

To summarise, in procedural tasks, the arrangement of information in space can be used in three ways. First, space can reflect the visuo-spatial arrangement of elements as present in the real world. Second, space can be mapped onto time, by illustrating different states of the elements. Third, space can be used to present actions performed on the elements.

1.5.4 Linking theoretical accounts to task type

From the descriptions of these task types, it can be concluded that there are clear differences amongst multimedia materials for conceptual, causal, and procedural tasks. These differences between conceptual, causal, and procedural tasks pertain, among other things, to how important space is for understanding the learning material. In conceptual tasks, spatial arrangement is used metaphorically to represent non-spatial relationships amongst objects. This implies that spatial arrangements in conceptual tasks can be changed (e.g., mirrored horizontally or vertically) without losing its meaning, and that the critical information about relationships amongst objects should be obtained from the text. In contrast, in causal and procedural tasks, the spatial arrangement of elements in the picture, map the spatial arrangement and order in the real world. This implies that the spatial arrangement of objects cannot be changed without changing its meaning and that critical information about relationships amongst objects can be obtained from the picture. As the critical information about relationships amongst objects should be obtained from the text in conceptual tasks and from the picture in causal and procedural tasks, it can be argued that in causal and procedural tasks, visuo-spatial information is more important for understanding the learning material.

As according to the account of computational offloading, pictures facilitate information processing during learning especially when visuo-spatial information is processed, it can be argued that more computational offloading is possible when more visuo-spatial information is processed: with more visuo-spatial information, imagery is more necessary for understanding when learning from text only, implying that pictures are more computational offloading. Therefore, based on the account of computational offloading, it can be predicted that pictures are beneficial in conceptual, causal, and procedural tasks (i.e., a multimedia effect is expected for all task types), but that this beneficial effect is larger in causal and procedural tasks than in conceptual tasks (i.e., the size of the multimedia effect is expected to be larger in causal and procedural tasks than in conceptual tasks).

In contrast, CTML predicts that pictures are beneficial to learning, because two codes are stored in memory, but does not allow assuming that studying tasks that differ in how important visuo-spatial information is for understanding the learning material, will lead to differences in dual coding. Therefore, based on CTML, it can be assumed that pictures are beneficial in conceptual, causal, and procedural tasks (i.e., a multimedia effect is expected for all task types), and that this beneficial effect is similar across task types (i.e., the size of the multimedia effect is expected to be the same for all task types).

As described before, the three important differences between CTML and the account of computational offloading pertain to their predictions concerning whether the amount of visuo-spatial information required for understanding the learning material, affects multimedia learning (cf. concept of task type), and whether information processing in working memory differs during multimedia learning (cf. concept of working memory). As the concept of task type has been described in detail in section 1.5, the concept of working memory is introduced in more detail in section 1.6. Here, the most commonly used model concerning working memory in the field of multimedia learning, namely Baddeley's working memory model (1986; Section 1.6.1) and how working memory involvement can be assessed (Section 1.6.2) is introduced first. After that, an overview of research that has been published concerning information processing in working memory in multimedia learning is provided (Section 1.6.3). Finally, the discrepancy between empirical results and theoretical predictions are described (Section 1.6.4) and hypotheses concerning the beneficial effect of pictures and the role of working memory during learning are proposed (Section 1.6.5).

1.6 Working Memory Involvement during Multimedia Learning

1.6.1 Structure of working memory

Baddeley's working memory model (1986) posits that working memory consists of multiple subsystems: the central executive, the phonological loop, and the visuo-spatial sketchpad. The central executive is assumed to be responsible for a) coordinating and monitoring the phonological loop and the visuo-spatial sketchpad, b) linking these subsystems to long-term memory, c) switching attention between tasks, d) allocate attention to (parts of the) stimuli, e) assigning information processing to one of the subsystems, and f) updating and controlling working memory contents (Baddeley 1996; Smith & Jonides 1999). As the central executive was assumed to have no storage capacity, it was difficult to explain all processes for which the central executive was held responsible. Therefore, newer versions of the model also include the episodic buffer (Baddeley, 2000).

The episodic buffer allows temporary storage of information from both the phonological loop and the visuo-spatial sketchpad, so that this information can be combined and integrated with prior knowledge (Baddeley, 2000). The phonological loop is responsible for processing and storing verbal information, both in the form of spoken and written words (see also Footnote 1). Therefore, the phonological loop is assumed to be involved during text processing when learning with multimedia. The visual part of the visuo-spatial sketchpad is assumed to be responsible for processing visual characteristics of objects, such as shape and colour (Logie, 1995), whereas the spatial part is assumed to be responsible for processing spatial and relational information, such as position, and movement control, such as eye movements (Lawrence, Myerson, Oonk, & Abrams, 2001; Logie 1995; Logie & Marchetti 1991). Therefore, the visuo-spatial sketchpad is assumed to be involved during picture processing, motor control, and kinaesthetic information processing. This distinction between the phonological loop and visuo-spatial sketchpad for processing verbal versus visuo-spatial information respectively has been adopted by CTML. However, more recent working memory research has shown that this distinction is too straightforward (Gyselinck et al., 2007) by showing that text content addressing visuo-spatial information is processed in the visuo-spatial sketchpad instead of the phonological loop. This is in line with the idea that visuo-spatial information processing such as imagery occurs in the visuo-spatial sketchpad. Therefore, the visuo-spatial sketchpad is not only assumed to processes pictures, but is also assumed to be involved during mental imagery (Baddeley, 2012; Smyth & Pendleton, 1990). This assumption is a prerequisite for the account of computational offloading, as visuo-spatial perception is assumed to enable computational offloading in the visuo-spatial sketchpad compared to the more effortful and error prone process of imagery (see Section 1.3.1).

1.6.2 Assessing working memory involvement

The working memory subsystems are said to have a limited capacity to process information, which implies that only a limited amount of information can be processed at the same time (Baddeley, 1999). When two tasks involve information processing in different subsystems (e.g., studying a picture and listening to a spoken text) this information can be processed relatively independently from each other. However, when two tasks require information processing in the same subsystem (e.g., reading a written text and listening to a different spoken text), it is probable that not all information can be processed at the same time, which leads to interference between the two tasks and thereby impaired performance on one or both tasks (Baddeley, 1999).

To investigate information processing in working memory during learning, the dual task approach can be used. This approach allows measuring the load of working memory subsystems when performing the primary task, in this case multimedia learning. The underlying principle of performing the dual task is that this task requires the resources of one of the working memory subsystems and therefore causes interference with the learning task (Andrade, 2001). If learning outcomes are affected by performing this dual task, it can be concluded that the working memory subsystem that was loaded by the dual task was involved during learning (Baddeley, 1992).

The involvement of working memory subsystems during learning can be assessed using a secondary task and a preload task (Andrade, 2001; Cocchini, Logie, della Sala, MacPherson, & Baddeley, 2002). The most commonly used dual task is the secondary task (see Schüler, Scheiter, & van Genuchten, 2011 for a review), which involves performing a secondary task during the primary task. Performing such tasks during learning is assumed to require resources in one of the working memory subsystems (Andrade, 2001). The involvement of working memory subsystems during learning can also be measured using a preload task, which involves memorising information before performing the primary task and judging whether this information is identical to the information presented after performing the primary task (cf. Cocchini et al., 2002). Irrespective of the type of dual task, stronger interference between the dual task and information processing implies stronger involvement of the working memory subsystems during performing the primary task. Such interference can, for example, be measured by comparing learning outcomes of participants who learned with dual task and participants who learned without dual task. If there is interference, participants in the dual task condition will have lower learning outcomes than participants who learned without dual task.

To measure the involvement of the central executive, random generation secondary tasks can be used. In these tasks, random sequences are generated by the participant during learning by naming letters or numbers or by tapping keys or pedals in a random order (e.g., Baddeley, Emslie, Kolodny, & Duncan, 1998). It is argued that these tasks not only require information processing in one of the subsystems to, for example, process letters, but also additional resources in the central executive to prevent uttering common sequences, such as ‘123’ and ‘abc’. As always one of the working memory subsystems is involved, at least two random generation tasks should be used: one loading the phonological loop and one loading the visuo-spatial sketchpad. To my knowledge, no preload task exists to measure the involvement of the central executive. Also, to my knowledge, no dual task or preload task has yet been developed to measure the involvement of the episodic buffer.

To measure the involvement of the phonological loop using a secondary task, the articulatory suppression task can be used. In this task, syllables, words, or numbers are repeated in a particular order at a particular pace during learning (Murray, 1967). The purpose of this task is to disturb rehearsal and recoding of verbal information during the learning task (Gathercole & Baddeley, 1993). When using a preload task, a verbal stimulus, such as a list of words, can be presented before the learning task, which has to be recalled after learning (e.g., Cocchini et al., 2002).

To measure the involvement of the visuo-spatial sketchpad using a secondary task, the finger or foot tapping task can be used (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Miyake, Emerson, Padilla, & Ahn, 2004). These tasks interfere with visuo-spatial information processing, because spatial motor tasks require motor control, which, like the processing of visuo-spatial information from text or pictures, is handled by the visuo-spatial sketchpad (Baddeley, 1986). When using a preload task, a visuo-spatial stimulus, for example a matrix partly filled with dots, is presented before learning and has to be recalled after learning (e.g., Kruley, Sciama, & Glenberg, 1994).

In previous research, the dual approach has been used to investigate the role of working memory during multimedia learning. The results of these studies concerning the involvement of the working memory subsystems are described next.

1.6.3 The role of working memory during multimedia learning

The assumption of limited capacity of the working memory subsystems has been adopted by CTML (Mayer, 2005b), which is used within CTML as the grounding for the main argument that certain types of learning materials use working memory resources more efficiently than other types of learning materials. Also, the assumption of distinguishing between the phonological loop as the subsystem responsible for text processing and the visuo-spatial sketchpad as the subsystem responsible for picture processing is adopted by CTML (Mayer, 2005b), although this assumption is not in line with contemporary research (Gyselinck et al., 2007). However, the role of the central executive or the episodic buffer is not made explicit, although CTML assumes that verbal and pictorial information are integrated with each other and with prior knowledge. Research investing the role of the central executive, phonological loop, and visuo-spatial sketchpad using the dual task approach is described in the following.

1.6.3.1 Role of the central executive

The role of the central executive during multimedia learning has to my knowledge only been investigated by Brunyé et al. (2006). They examined working memory processes using secondary tasks while studying procedural tasks that focused on how to assemble Kinder Egg™ toys in text-only, multimedia (i.e., text and picture), and picture-only format. Their tasks conveyed procedures by describing and depicting states, but did not describe interactions between objects and body parts of the person performing the actions. Participants performed either a random generation articulatory task or a random generation tapping task to load the central executive, an articulatory suppression task to load the phonological loop, or a spatial tapping task to load the visuo-spatial sketchpad. Learning outcome tasks involved verifying whether two sequence steps appeared in the correct temporal order, describing each sequence, and recalling in which format the procedure was studied. Their results showed that performing a random generation dual task interfered more with learning when studying text and pictures than studying text only or picture only, which indicates that the central executive is involved in multimedia learning.

1.6.3.2 The role of the phonological loop

The role of the phonological loop during multimedia learning has been investigated several times. The results of the study by Brunyé et al. (2006) described earlier showed that the articulatory suppression secondary task interfered with text processing, but not with picture processing. This result indicates that the phonological loop was involved only in text processing.

Kruley et al. (1994) examined working memory processes while studying causal tasks concerning biology and physics, using preload dual tasks. They presented 32 tasks in text-only or multimedia format in a within-subjects design. The pictures represented the structural relationships between the parts of the object described in the text. The learning outcome task involved three multiple-choice questions, which tested the participants' comprehension of the material. They investigated the involvement of the phonological loop using a verbal preload task and found that this preload task impaired performance in both the text-only and text-picture condition, indicating that the phonological loop was involved in text processing. In contrast, the presentation of a picture did not influence performance when using a verbal preload task, indicating that the phonological loop was not involved in picture processing.

Gyselinck, Ehrlich, Cornoldi, de Beni, and Dubois (2000) also examined working memory processes using preload dual tasks while studying causal tasks that described basic notions of physics. They presented six tasks in text-only or multimedia format in a within-subjects design. The pictures represented the elements mentioned in the sentence and the causal and temporal relationships between them. Learning outcomes tasks involved testing factual information that was explicitly given in the text and involved testing students' ability to draw inferences from several sentences. They could not confirm the involvement of the phonological loop as found by Brunyé et al. (2006) and Kruley et al. (1994).

Finally, Gyselinck, Cornoldi, Dubois, de Beni, and Ehrlich (2002) used the same learning material and learning outcome measures, but used secondary tasks. Their results confirmed the results from Brunyé et al. (2006) and Kruley et al. (1994) that the phonological loop was involved in text processing, but not in picture processing: the articulatory suppression task impaired learning in both conditions, without affecting the advantage of picture presentation.

1.6.3.3 The role of the visuo-spatial sketchpad

The aforementioned studies also invested the role of the visuo-spatial sketchpad. The results of Brunyé et al. (2006) showed that the spatial tapping secondary task interfered only with picture processing, which indicates that the visuo-spatial sketchpad was involved in picture processing, but not in text processing.

Kruley et al. (1994) investigated the involvement of the visuo-spatial sketchpad using a visuo-spatial preload task. They found that their multimedia learning tasks loaded the visuo-spatial sketchpad, because when only spoken texts were presented, there was no interference with the visuo-spatial preload task, whereas there was interference when both text and pictures were presented. These results show that the visuo-spatial sketchpad was involved in processing both picture and matrices, but not in text processing.

Gyselinck et al. (2000) could not confirm the involvement of the visuo-spatial sketchpad as found by Kruley et al. (1994). However, the results Gyselinck et al. (2002) confirmed the results from Kruley et al. (1994) that the visuo-spatial sketchpad is involved in picture processing, but not in text processing: the spatial tapping task impaired learning when text and pictures were presented, but not when text only was presented.

1.6.4 Discrepancy between computational offloading account, recent working memory research and empirical results

To summarize, the results of Brunyé et al. (2006), Kruley et al. (1994), and Gyselinck et al. (2002) show that the phonological loop was involved in processing text during learning — irrespective of whether learning material was presented as text and pictures or text only — but not in picture processing, and that the visuo-spatial sketchpad was involved in picture processing, but not in text processing. Only Brunyé et al. (2006) investigated the involvement of the central executive, confirming the assumption that the central executive is involved when multimodal information is presented.

The results of Brunyé et al. (2006), Kruley et al. (1994), and Gyselinck et al. (2002) are surprising from the computational offloading point of view and from contemporary working memory research concerning the involvement of the visuo-spatial sketchpad during text processing. As described before, when contrasting imagery and picture processing, it seems to require more mental effort to generate a mental image (cf. Kirsh & Maglio, 1994), to interpret text (Larkin & Simon, 1987), to recall mental images, and to apply operations to these images (Kosslyn, 1999) than to perceive the actual picture (Larkin & Simon, 1987). Therefore, based on the idea that pictures are computational offloading, the visuo-spatial sketchpad would expected to be more — instead of less — involved when processing visuo-spatial information when learning with text only than when learning with text and pictures.

A possible explanation for this discrepancy is that, although imagery may play an important role during mental model construction from text only, learners in the text-only conditions did not apply imagery to understand and memorise the text. Instead, they might have used verbal encoding strategies, which require information processing in the phonological loop (cf. Mayer & Sims, 1994). Empirical research has shown that learners indeed do not always spontaneously apply imagery during learning; rather, they need to be explicitly instructed to apply imagery during learning. Previous research on mental imagery in causal tasks has shown that instructing participants to apply imagery enhanced learning for these tasks compared to a control condition in which participants were not instructed to apply imagery (e.g., Hegarty, Kriz, & Cate, 2003a; Hegarty, Narayanan, & Freitas, 2003b). However, in studies using procedural tasks conveying actions, instructing participants to imagine the motor actions did not enhance learning (e.g., golf putting [Woolfolk, Murphy, Gottesfeld, & Aitken, 1985], tennis serve [Weinberg, Gould, Jackson, & Barnes, 1980;

Epstein, 1980). Taken together, this could imply that participants often fail to spontaneously engage in imagery when studying causal tasks and therefore benefit from imagery instructions, whereas participants spontaneously engage in imagery when studying procedural tasks that convey actions and therefore do not benefit from imagery instructions. If imagery is indeed not applied when learning with text only, no load on the visuo-spatial sketchpad and no interference with the dual task is evoked. If only participants who learned with multimedia processed visuo-spatial information, this could explain the interference between the dual task and learning with pictures. Based on this postulation, it can be argued that when investigating working memory involvement during learning with text only versus text and pictures, procedural tasks that convey actions should be considered.

1.6.5 Linking theoretical accounts to working memory

When combining this description of the concept of working memory with CTML and the account of computational offloading, different predictions regarding the involvement of working memory during learning with multimedia can be derived. These predictions refer to 1) working memory involvement as a function of task type and 2) the role of working memory when learning from text only versus text and pictures. These predictions are addressed in two experiments.

When considering the predictions concerning working memory involvement as a function of task type, CTML would predict that all working memory subsystems are involved when learning with text and pictures. Also, as CTML predicts no difference in the size of the multimedia effect between task types (see Section 1.5.4), CTML would also predict no differences in the roles of these subsystems between task types. In case of the visuo-spatial sketchpad, CTML does not predict differences between task types, because CTML does not consider recent findings that show visuo-spatial sketchpad involvement during processing of visuo-spatial text content. In contrast, the account of computational offloading would predict differences in information processing between task types, as pictures are assumed to be computational offloading especially when visuo-spatial information is important for understanding the learning task. The account of computational offloading would argue that the visuo-spatial sketchpad is more involved in causal and procedural tasks, as mental imagery is more likely to be applied when tasks contain more visuo-spatial information and visuo-spatial information is more important for understanding the learning material, than in conceptual tasks. The account of computational offloading does not make predictions concerning the involvement of the central executive and phonological loop. As existing studies do not allow any firm conclusions concerning the role of working memory in different task types, because their learning tasks varied greatly, the role of working memory is reassessed within a single study (i.e., Experiment 2), in which task-unrelated characteristics and learning outcome measures for each task type were controlled.

When considering the predictions concerning the role of working memory when learning from text versus text and pictures, CTML would predict that picture processing requires more visuo-spatial information processing than text processing, as pictures are assumed to be processed in the visuo-spatial sketchpad whereas text is not. In contrast, the account of computational offloading would predict that the visuo-spatial sketchpad is more involved when learning with text only than when learning with text and pictures, as imagery is assumed to require more cognitive resources than visuo-spatial perception.

1.7 Hypotheses

In the previous sections of this introduction, the multimedia effect and two theoretical accounts that explain the multimedia effect have been discussed. To summarise, the Cognitive Theory of Multimedia Learning (CTML) predicts that pictures are beneficial to learning, because two codes are stored in memory, whereas the account of computational offloading predicts that pictures are beneficial to learning, because they facilitate information processing in working memory. Three important differences between CTML and the account of computational offloading were discussed, which pertain to their predictions concerning whether the amount of visuo-spatial information required for understanding the learning material affects multimedia learning (cf. concept of task type), and whether information processing in working memory differs during multimedia learning (cf. concept of working memory). Both theoretical accounts make different predictions concerning 1) the size of the multimedia effect in different task types, 2) working memory involvement in different task types, and 3) the role of the visuo-spatial sketchpad in working memory when learning with text only or with multimedia. In this dissertation, three experiments are described that aim at testing these predictions, to contrast both theoretical accounts and obtain insight into why pictures are beneficial to learning.

In Experiment 1, the effect of the importance of visuo-spatial information and thus of task type for the multimedia effect is investigated. CTML would predict that there will be a multimedia effect for conceptual, causal, and procedural tasks, and that the size of this multimedia effect is the same for all task types, as dual coding of information when learning with multimedia should occur irrespective of task type. The account of computational offloading would also predict a multimedia effect for conceptual, causal, and procedural tasks. However, in contrast to CTML, it would predict that this multimedia effect is larger in causal and procedural tasks than in conceptual tasks, as visuo-spatial information seems to be more important for understanding in causal and procedural tasks than conceptual tasks.

In Experiment 2, the effect of the importance of visuo-spatial information on working memory involvement is investigated. CTML would predict that the central executive, the phonological loop, and the visuo-spatial sketchpad are involved when learning with multimedia, and that there are no differences between conceptual, causal, and procedural tasks in working memory involvement. In contrast, the account of computational offloading would predict that the involvement of the visuo-spatial sketchpad increases when visuo-spatial information becomes more important for understanding the learning task. Therefore, it would predict that the involvement of the visuo-spatial sketchpad is higher in causal and procedural tasks than in conceptual tasks. No differences between conceptual, causal, and procedural tasks would be predicted concerning the involvement of the central executive and phonological loop.

In Experiment 3, the account of computational offloading is tested by comparing the involvement of the visuo-spatial sketchpad when learning with text only or with multimedia, in procedural tasks that are assumed to trigger imagery. CTML would predict that the visuo-spatial sketchpad is more involved when learning with multimedia than when learning with text only, as the visuo-spatial sketchpad is assumed to be involved only in picture processing. In contrast, the account of computational offloading would predict that the visuo-spatial sketchpad is more involved when learning with text only than when learning with multimedia, as imagery requires more mental resources than visuo-spatial perception.

2 Experiment 1[2]

Existing studies do not allow any firm conclusions concerning the explanation of the beneficial effect of pictures in multimedia learning tasks that differ in how important visuo-spatial information is for understanding the task, because they vary on too many task-unrelated dimensions and on how learning outcomes are gauged. Accordingly, the purpose of the first experiment was to investigate the multimedia effect for conceptual, causal, and procedural tasks within a single study, while controlling for task-unrelated characteristics and using the same learning outcome measures for each task type. CTML, of which the explanation of the multimedia effect is based on dual coding, would predict a multimedia effect for conceptual, causal, and procedural tasks, and that the size of this multimedia effect is the same for all task types. The account of computational offloading, which focuses on facilitated information processing during learning, would also predict a multimedia effect for conceptual, causal, and procedural tasks. However, in contrast to CTML, it would predict that this multimedia effect is larger in causal and procedural tasks than in conceptual tasks as visuo-spatial information is more important for understand in causal and procedural tasks than in conceptual tasks.

2.1 Method

2.1.1 Participants and design

A 2x3x3 design was used, with presentation type as the between-subject variable, and task type and information source as the within-subject variables. Sixty-five university students (49 female and 16 male; *M* = 23.54 years, *SD* = 3.34), of which approximately half of the students were psychology students and the other half had varied university backgrounds, were randomly assigned to one of two presentation types. Depending on presentation type, students received just a text (text-only condition; *n* = 32) or a text accompanied by a picture (multimedia condition; *n* = 33). All students studied materials of the three task types — conceptual, procedural, and causal tasks. For each task, three sources of information were

[2] This experiment has been published as: van Genuchten, E., Scheiter, K., & Schüler, A. (2012). Examining learning from text and pictures for different task types: Does the multimedia effect differ for conceptual, causal, and procedural tasks? *Computers in Human Behavior*, *28*, 2209-2218.

defined — information given in text only, in text and picture, or in picture only. Levie and Lentz (1982) suggest that greater clarity concerning the beneficial effect of pictures can be obtained by distinguishing between information sources. Students received either payment or course credit.

2.1.2 Materials and measures

2.1.2.1 Learning tasks

Nine learning tasks were developed of which three were conceptual, three causal, and three procedural. Each student studied six out of these nine learning tasks, two from each type. All tasks were set in a fictitious country to ensure that prior content knowledge was limited. By using fictitious material, it was also possible to make these tasks as comparable as possible (e.g., number of propositions, degree of overlap between information given in the text and picture, structure of the text, style of the drawings), and to create unrelated tasks, in so that information from one task did not influence performance on other tasks. Each task contained 1/2 A4-page written text, which was accompanied by one composite picture in the multimedia condition. A propositional analysis (Bovair & Kieras, 1985) showed that the total number of propositions (i.e., in text and in pictures) in conceptual tasks was on average 125, in causal tasks 143, and in procedural tasks 132. For all tasks, the number of propositions only mentioned in the text ranged from 68 to 77, mentioned only in the picture from 28 to 66, and mentioned in both text and picture (i.e., overlap) from 10 to 18. The tasks were presented on a computer screen.

The conceptual tasks focused on concepts and their relationships. Conceptual tasks used in this experiment concerned the relationships between inhabitants of a village (see Figure 2), the relationships between tribes and their trade connections, and an animals' food hierarchy. The causal tasks focused on cause-and-effect chains. Causal tasks used in this experiment concerned a machine to notify a village's inhabitants that eggs should be collected (see Figure 3), a machine to send a message to one of four gods, and a machine to warn for danger. The procedural tasks focused on the temporal and spatial relationships between states. Procedural tasks used in this experiment concerned how to bend a tree to obtain drinkable juice (see Figure 4), how to build a tent, and how to sew a headscarf.

2.1.2.2 Post-tests

To test how much knowledge students had obtained, verbal and pictorial test items were used. Four dependent variables were constructed from these items. For three types of tests (free recall, recall verification accuracy, and transfer verification accuracy) a distinction was made between information sources (information given in text only, in picture only, or in both text and picture; within-subject variable). The fourth dependent variable was integration verification accuracy, for which no distinction between information sources was made: in these questions, information from the text had to be integrated with information from the picture, implying that both sources were needed to answer these questions correctly.

In the free recall test, students wrote down and drew everything that they remembered. I created a propositional scoring system for each task using the propositional analysis that I also used to count the total number of propositions. In this scoring system, it was ensured that one point could be assigned to each relevant piece of information. If the same piece of information was recalled twice (e.g., once in the text and once in the drawing) only one point was awarded. To assess the inter-rater reliability, two raters coded 20% of all texts and drawings. The Cohen's kappa was .71, which is considered good (Landis & Koch, 1977). The remaining 80% was scored by a single rater only. Students could score between 0 and 1 proportion recalled.

In the recall verification tests, students answered yes/no recall verification questions. For each task type, they answered nine verbal questions from three information sources. Three questions addressed information only given in the text (e.g., The son of the baker has a girlfriend), three addressed information only given in the picture (e.g., Hexagon's name is Yellow), and three addressed information given in both text and picture (e.g., Blue Circle has five children). They also answered six pictorial verification questions for each task from two information sources. Three addressed information only given in the picture (see Figure 6a) and three addressed information given in both text and picture (see Figure 6b). One point was assigned to each correct answer. From these questions, three dependent variables (from the three information sources) were constructed. Therefore in each task, students scored for information from text only between 0 and 3 (i.e., for three verbal items), whereas for information from text-picture and picture only, students scored between 0 and 6 (i.e., for three verbal and three pictorial items).

In the transfer verification test, students answered yes/no transfer verification questions. For each task type, they answered nine verbal questions of which three addressed information only given in the text (e.g., If the mayor moves to the same street as Orange, then he lives in the same street as the baker), three addressed information only given in the picture (e.g., If children are born, they never have the same colour as their parents), and three addressed information given in both text and picture (e.g., If Red would be a circle, then all children would be circles). One point was assigned to each correct answer. From these questions, three dependent variables (from the three information sources) were constructed. In each task and for each variable, students scored between 0 and 3.

In the integration verification test, students answered yes/no integration verification questions. They answered three questions per task. One point was assigned to each correct answer. To be able to answer these questions, information from the text and information from the picture needed to be integrated. In each task, students scored between 0 and 3. An example is given in Figure 6c.

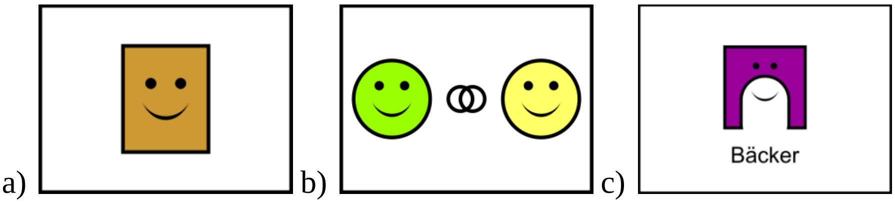

Figure 6. Example of pictorial verification questions (a and b) and an integration verification question (c).

For the recall verification accuracy, transfer verification accuracy, and integration verification accuracy questions, it was ensured that half of these statements were correct (requiring a ‘yes’ answer by pressing the ‘yes’-button on the button box) and half of the statements incorrect (requiring a ‘no’ answer by pressing the ‘no’-button). This implies that the chance of answering a question correct by guessing was .50. All verification questions were presented on a computer screen.

2.1.3 Procedure

Before starting the experiment, students gave their informed consent, and filled in a demographic questionnaire. Then, students performed two blocks. Each block consisted of three learning tasks (one of each task type) and post-tests for each task. Each student studied six tasks in total, whereby the selection of tasks from each task type was random. The post-tests were conducted in the same order in which the material was learned. For the free recall task, the order of drawing and writing was counterbalanced. Students were allowed to write for a maximum of five minutes and to draw for a maximum of five minutes for each task. Before students answered the verification questions of the first learning task, they saw an example of a verbal, a pictorial, and an integration question. Then participants answered all verbal questions, after that all pictorial items, and after that all integration questions. The verbal, pictorial, and integration verification questions were presented in random order within each type of question. After the first block, students had a short break in which the student and the experimenter played the game Mikado™. Both learning and post-test phase (except for free recall) were learner-paced. The duration of one session was between 2 and 2.5 hours.

2.1.4 Data analysis

To be able to investigate the multimedia effect for conceptual, causal, and procedural tasks, for different information sources when considering different types of learning outcomes, four mixed ANOVAs were conducted. The dependent variables for these analyses were free recall, recall verification accuracy, transfer verification accuracy, and integration verification accuracy. The between-subjects factor was presentation type. As within-subjects factor, task type (conceptual, causal, procedural) was included. For free recall, recall verification accuracy, and transfer verification accuracy, information source (information given in text only, text and picture, or picture only) was included as a second within-subjects factor. The main effects of task type and of information source, and the interaction between task type and information source are not reported, due to the lack of meaningful interpretations and the lack of theoretical relevance (these effects merely say something about the relative difficulty of the learning materials and the post-test questions). All statistical assumptions for the reported tests were met. All effects are reported as significant at $p < .05$. Partial eta-squared was used as effect size, which reflects the amount of variance that is explained by a model when other non-error sources of variance are partialled out (Cohen, 1973). Here, .01, .06, and .14 correspond to small, medium, and large effect sizes respectively (Cohen, 1988). For significant overall effects, Bonferroni post-hoc tests were conducted. Also, in the case of a significant interaction between presentation type and task type, contrasts were used to determine how the multimedia effect differed as a function of task type. In particular, these tests assessed how the size of the difference between the text-only and the multimedia condition (i.e., the size of the multimedia effect) varied between conceptual, causal, and procedural tasks.

2.2 Results

Descriptive statistics are provided in Table 2. An overview of the results for the multimedia effects is provided in Table 3.

2.2.1 Free recall

The results of the mixed ANOVA with free recall as dependent variable showed a main effect of presentation type ($F(1, 63) = 66.66$, $p < .001$, $\eta_p^2 = .51$) implying that students in the multimedia condition ($M = .50$, $SD = .11$) learned more than students in the text-only condition ($M = .33$, $SD = .12$). This means that there was a multimedia effect. However, it should be noted that this main effect was qualified by several interactions between information source, presentation type, and task type.

First, there was a significant interaction between task type and presentation type ($F(2, 62) = 6.03$, $p < .01$, $\eta_p^2 = .16$). Post-hoc comparisons showed a multimedia effect in all task types; however, the size of the multimedia effect differed between task types. Helmert contrasts showed that the difference between the text-only and the multimedia condition (i.e., the multimedia effect) was smallest in conceptual and causal tasks, and largest in procedural tasks (conceptual vs. causal and procedural tasks: $F(1, 63) = .78$, $p = .38$; causal vs. procedural: $F(1, 63) = 12.11$, $p < .01$; conceptual: text-only: $M = .34$, $SD = .13$, multimedia: $M = .49$, $SD = .13$, $p < .001$, $\eta_p^2 = .32$; causal: text-only: $M = .37$, $SD = .12$, multimedia: $M = .52$, $SD = .07$, $p < .001$, $\eta_p^2 = .41$; procedural: text-only: $M = .29$, $SD = .12$, multimedia: $M = .51$, $SD = .12$, $p < .001$, $\eta_p^2 = .54$).

Second, the interaction between information source and presentation type was significant ($F(2, 62) = 139.43$, $p < .001$, $\eta_p^2 = .82$). Post-hoc comparisons showed that there was no multimedia effect for text questions ($p = .76$, $\eta_p^2 < .01$; text-only: $M = .21$, $SD = .08$; multimedia: $M = .21$, $SD = .06$). However, there was a multimedia effect for text-picture questions ($p < .001$, $\eta_p^2 = .40$; text-only: $M = .60$, $SD = .17$; multimedia: $M = .79$, $SD = .13$), and a larger multimedia effect for picture questions ($p < .001$, $\eta_p^2 = .72$; text-only: $M = .20$, *SD* = .10; multimedia: *M* = .51, *SD* = .13). Polynomial contrasts showed that the differences between the text-only and multimedia condition (i.e., the size of the multimedia effect) for text, text-picture, and picture questions increased linearly ($F(1, 63) = 283.21$, $p < .001$).

Finally, the results showed a significant three-way interaction between information source, task type, and presentation type ($F(4, 60) = 2.90$, $p = .03$, $\eta_p^2 = .16$; see Figure 7). Post-hoc comparisons showed, for conceptual tasks, no multimedia effect for text questions ($p = .72$, $\eta_p^2 < .01$), a multimedia effect for text-picture questions ($p < .001$, $\eta_p^2 = .21$), and a larger multimedia effect for picture questions ($p < .001$, $\eta_p^2 = .55$). Similarly, for causal tasks, the results showed no multimedia effect for text questions ($p = .61$, $\eta_p^2 < .01$), a multimedia effect for text-picture questions ($p < .001$, $\eta_p^2 = .23$), and a larger multimedia effect for picture questions ($p < .001$, $\eta_p^2 = .72$). For procedural tasks, the results showed a multimedia effect for all sources of information. This effect was smallest for text questions ($p = .04$, $\eta_p^2 = .06$), larger for text-picture questions ($p < .001$, $\eta_p^2 = .46$), and largest for picture questions ($p < .001$, $\eta_p^2 = .64$). Helmert and polynomial contrasts yielded a significant quadratic trend indicating that the difference between the text-only and the multimedia condition (i.e., the size of the multimedia effect) for text-picture questions in procedural tasks was larger than for text-picture questions in conceptual and procedural tasks (conceptual vs. causal and procedural: $F < 1$; causal vs. procedural: $F(1, 63) = 11.83$, $p < .01$).

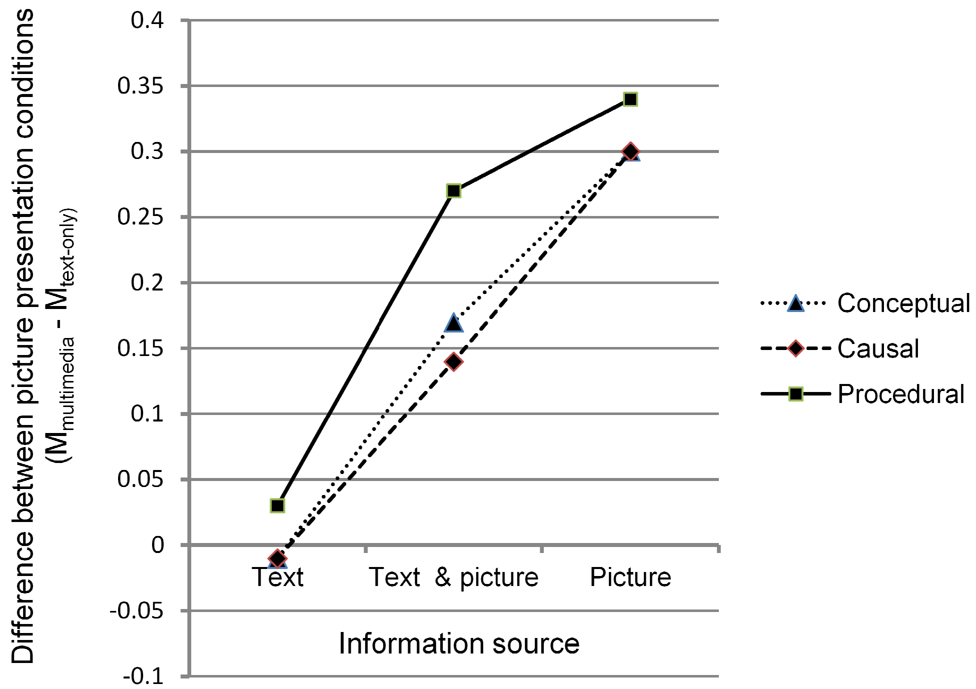

Figure 7. The difference in means between the multimedia and the text-only condition for free recall as a function of information source and task type.

2.2.2 Recall verification accuracy

The results of the mixed ANOVA with recall verification accuracy as dependent variable showed a main effect of presentation type ($F(1, 63) = 19.60$, $p < .001$, $\eta_p^2 = .24$) implying that students in the multimedia condition ($M = .80$, $SD = .13$) learned more than students in the text-only condition ($M = .73$, $SD = .13$). This means that there was a multimedia effect. Again, this main effect was qualified by several interactions between information source, presentation type, and task type.

First, there was a significant interaction between task type and presentation type ($F(2, 62) = 8.11$, $p < .01$, $\eta_p^2 = .21$). Post-hoc comparisons only showed a multimedia effect in procedural tasks (conceptual: text-only: $M = .75$, $SD = .13$, multimedia: $M = .79$, $SD = .15$, $p = .08$, $\eta_p^2 = .05$; causal: text-only: $M = .76$, $SD = .13$, multimedia: $M = .78$, $SD = .13$, $p = .27$, $\eta_p^2 = .02$; procedural: text-only: $M = .69$, $SD = .13$, multimedia: $M = .82$, $SD = .12$, $p < .001$, $\eta_p^2 = .42$).

Second, the interaction between information source and presentation type was significant ($F(2, 62) = 20.01$, $p < .001$, $\eta_p^2 = .39$). Post-hoc comparisons showed that there was no multimedia effect for text questions (*p* = .26, η_p^2 = .02; text-only: *M* = .87, *SD* = .14; multimedia: *M* = .84, *SD* = .14). However, there was a multimedia effect for text-picture questions (*p* < .01, η_p^2 = .14; text-only: *M* = .75, *SD* = .12; multimedia: *M* = .81, *SD* = .11), and a larger multimedia effect for picture questions (*p* < .001, η_p^2 = .51; text-only: *M* = .58, *SD* = .12; multimedia: *M* = .74, *SD* = .14). Polynomial contrasts showed that the difference between the text-only and the multimedia condition (i.e., the size of the multimedia effect) for text, text-picture, and picture questions increased linearly (*F*(1, 63) = 36.97, *p* < .001).

Finally, the results showed a significant three-way interaction between information source, task type, and presentation type (*F*(4, 60) = 3.31, *p* = .02, η_p^2 = .18; see Figure 8). Post-hoc comparisons showed, for conceptual tasks, only a multimedia effect for picture questions (text: *p* = .47, η_p^2 < .01; text-picture: *p* = .19, η_p^2 = .03, picture: *p* < .01, η_p^2 = .14). For causal tasks, the results showed no multimedia effect for text questions (*p* = .24, η_p^2 = .02), however did show a multimedia effect for text-picture questions (*p* = .04, η_p^2 = .06), and an equally large multimedia effect for picture questions (*p* = .03, η_p^2 = .07). Similarly, for procedural tasks, the results showed no multimedia effect for text questions (*p* = .97, η_p^2 < .001), however did show a multimedia effect for text-picture questions (*p* = .01, η_p^2 = .12), and a larger multimedia effect for picture questions (*p* < .001, η_p^2 = .59). The Helmert and linear contrasts showed that the difference between the text-only and the multimedia condition (i.e., the size of the multimedia effect) for text, text-picture, and picture questions increased in all task types, however that this increase was larger in procedural tasks than in causal tasks (conceptual vs. causal and procedural: *F* < 1; causal vs. procedural: *F*(1, 63) = 7.57, *p* = .01). Also, the Helmert and quadratic contrasts yielded a quadratic trend showing that the increase in differences between the text-only and the multimedia condition for text, text-picture and picture questions was stronger in procedural tasks than in causal tasks (conceptual vs. causal and procedural: *F* < 1; causal vs. procedural: *F*(1, 63) = 5.13, *p* = .03).

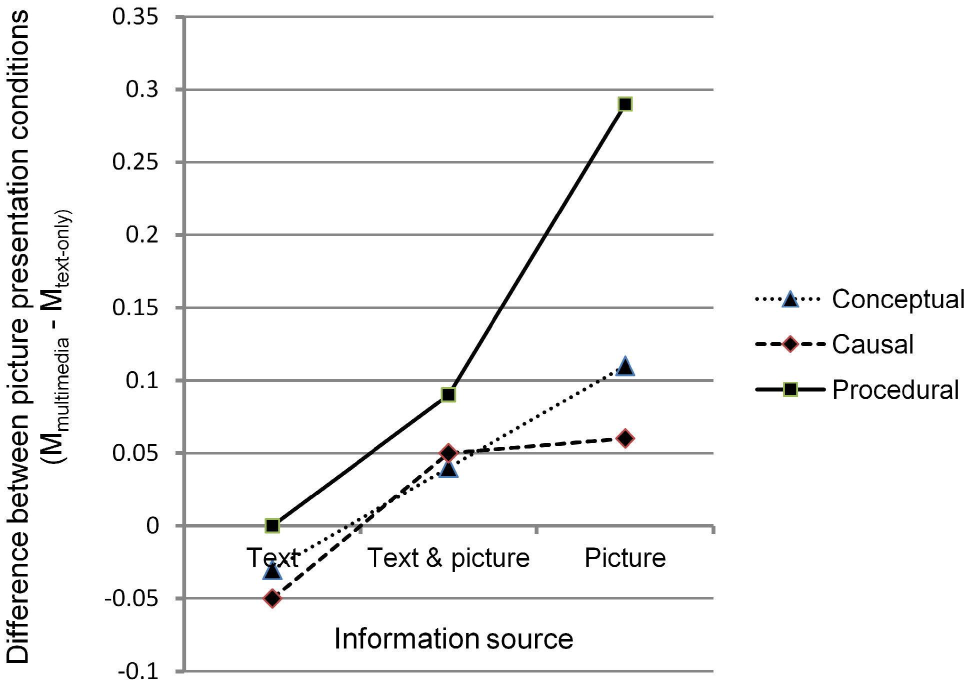

Figure 8. The difference in means between the multimedia and the text-only condition for recall verification accuracy as a function of information source and task type.

2.2.3 Transfer verification accuracy

The results of the mixed ANOVA with transfer verification accuracy as dependent variable showed no main effect of presentation type ($F(1, 63) = 3.18$, $p = .08$, $\eta_p^2 = .05$), implying that in general students in the multimedia condition ($M = .69$, $SD = .13$), did not learn better than students in the text-only condition ($M = .66$, $SD = .21$). This means that there was no multimedia effect.

There was also no significant interaction between task type and presentation type ($F(2, 62) = 1.84$, $p = .17$, $\eta_p^2 = .06$), implying that the differences between the presentation types were the same for all task types.

However, the interaction between information source and presentation type was significant ($F(2, 62) = 5.21$, $p = .01$, $\eta_p^2 = .14$). Post-hoc comparisons showed that there was no multimedia effect for text questions ($p = .26$, $\eta_p^2 = .02$; text-only: $M = .74$, $SD = .19$; multimedia: $M = .71$, $SD = .19$) and for text-picture questions ($p = .33$, $\eta_p^2 = .02$; text-only: $M = .62$, $SD = .19$; multimedia: $M = .65$, $SD = .17$). However, there was a multimedia effect for picture questions ($p < .01$, $\eta_p^2 = .13$; text-only: $M = .61$, $SD = .24$; multimedia: $M = .72$, $SD =$

.21). Polynomial contrasts showed that the difference between the text-only and the multimedia condition (i.e., the size of the multimedia effect) for text, text-picture, and picture questions increased linearly ($F(1, 63) = 10.58$, $p < .01$).

Finally, the results showed no significant three-way interaction between information source, task type, and presentation type ($F(4, 60) = 1.38$, $p = .25$, $\eta_p^2 = .08$), implying that the difference between the presentation types were the same in each task type for each information source.

Table 2

Means and Standard Deviations for Free Recall, Recall Verification Accuracy, and Transfer Verification Accuracy as a Function of Picture-presentation condition, Information Source, and Task Type

	Picture-presentation condition	Text-only			Multimedia		
	Information source	Text	Text-picture	Picture	Text	Text-picture	Picture
Dependent variable	Task type	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Free recall (proportion recalled)	Conceptual	.22 (.10)	.58 (.18)	.21 (.11)	.21 (.07)	.75 (.16)	.51 (.16)
	Causal	.21 (.08)	.71 (.17)	.20 (.09)	.20 (.05)	.85 (.07)	.50 (.09)
	Procedural	.19 (.06)	.51 (.16)	.18 (.11)	.22 (.05)	.78 (.14)	.52 (.14)
Recall verification accuracy (proportion correct)	Conceptual	.88 (.15)	.81 (.13)	.57 (.11)	.85 (.15)	.85 (.14)	.68 (.16)
	Causal	.85 (.15)	.78 (.12)	.65 (.11)	.80 (.16)	.83 (.08)	.71 (.13)

	Procedural	.87	.67	.53	.87	.76	.82
		(.14)	(.12)	(.13)	(.12)	(.12)	(.12)
Transfer	Conceptual	.68	.55	.44	.66	.60	.66
verification		(.22)	(.17)	(.26)	(.18)	(.16)	(.25)
accuracy	Causal	.80	.64	.67	.74	.65	.77
(proportion correct)		(.18)	(.21)	(.23)	(.21)	(.22)	(.20)
	Procedural	.74	.67	.71	.72	.70	.72
		(.19)	(.21)	(.23)	(.19)	(.13)	(.20)

2.2.4 Integration verification accuracy

In the final test, a mixed ANOVA was conducted with integration verification accuracy as dependent variable. The results showed a main effect of presentation type ($F(1, 63) = 10.19$, $p < .01$, $\eta_p^2 = .14$), implying that students in the multimedia condition ($M = .79$, $SD = .14$) performed better than students in the text-only condition ($M = .69$, $SD = .20$). This means that there was a multimedia effect.

However, the interaction between task type and presentation type was also significant ($F(2, 62) = 4.41$, $p = .02$, $\eta_p^2 = .13$). Post-hoc comparisons showed that there was a multimedia effect in conceptual tasks ($p < .01$, $\eta_p^2 = .15$; text-only: $M = .66$, $SD = .19$; multimedia: $M = .80$, $SD = .15$), and a multimedia effect in causal tasks ($p < .01$, $\eta_p^2 = .14$; text-only: $M = .71$, $SD = .22$; multimedia: $M = .85$, $SD = .13$). Surprisingly, there was no multimedia effect in procedural tasks ($p = .79$, $\eta_p^2 < .01$; text-only: $M = .71$, $SD = .19$; multimedia: $M = .72$, $SD = .15$). Helmert contrasts showed that the multimedia effect was equally large in conceptual and causal tasks, and smaller in procedural tasks (conceptual vs. causal and procedural: $F(1, 63) = 2.15$, $p = .15$; causal vs. procedural: $F(1, 63) = 6.11$, $p = .02$).

Table 3

Overview of Significance of the Multimedia Effect and the Effect Sizes (η_p^2) as a Function of Task Type and Information Source

		Free recall	Recall verification accuracy	Transfer verification accuracy	Integration verification accuracy
Text	Conceptual	-	-	-	
		< .01	.01	< .01	
	Causal	-	-	-	
		< .01	.02	.02	
	Procedural	+	-	-	
		.06	< .01	< .01	
Text-picture	Conceptual	+	-	-	+
		.21	.03	.02	.15

	Causal	+	+	-	+
		.23	.06	< .01	.14
	Procedural	+	+	-	-
		.46	.12	.01	< .01
Picture	Conceptual	+	+	+	
		.55	.14	.16	
	Causal	+	+	-	
		.72	.07	.05	
	Procedural	+	+	-	
		.64	.59	< .01	

2.3 Discussion

The aim of the first experiment was to investigate whether the beneficial effect of pictures differs between task types that differ in how important visuo-spatial information is for understanding the learning material. Therefore, the sizes of the differences between a text-only and a multimedia condition (i.e., the sizes of the multimedia effect) for conceptual, causal, and procedural tasks were compared. Four learning outcome measures (i.e., free recall, recall verification accuracy, transfer verification accuracy, and integration verification accuracy) and three information sources (i.e., information given in text only, in text and picture, or in picture only) were used. CTML would predict that the multimedia effect is equally large in these task types, whereas the account of computational offloading would predict that the multimedia effect is larger in procedural and causal than in conceptual tasks.

The results showed a multimedia effect for free recall, recall verification accuracy, and integration verification accuracy. However, in all cases, this main effect was qualified by interactions. For both free recall and recall verification accuracy, there was a three-way interaction between the multimedia effect, task type, and information source. In case of free recall, for conceptual tasks, no multimedia effect was found for text questions, a multimedia effect was found for text-picture questions, and a larger multimedia effect was found for picture questions. The same pattern was found for causal tasks. For procedural tasks, a multimedia effect was found for all information sources, which was smallest for text questions, larger for text-picture questions, and largest for picture questions. Irrespective of information source, the multimedia effect was largest in procedural tasks, and equally large in causal tasks and conceptual tasks.

In case of recall verification accuracy, a rather similar pattern was found. For conceptual tasks, a multimedia effect was found only for picture questions. For causal tasks, no multimedia effect was found for text questions, however an equally large multimedia effect was found for text-picture and picture questions. For procedural tasks, no multimedia effect was found for text questions, a multimedia effect was found for text-picture questions, and a larger multimedia effect was found for picture questions. Irrespective of information source, there was a multimedia effect only for procedural tasks.

Furthermore, for integration verification accuracy, there was a two-way interaction between the multimedia effect and task type (information source was not included as in these questions, because information from the text had to be integrated with information from the picture, implying that both sources were needed to answer these questions). Here, the opposite pattern was found: the multimedia effect was equally large in conceptual and causal tasks, and there was no multimedia effect in procedural tasks.

Finally, even though there was no multimedia effect for transfer verification questions, there was an interaction between the multimedia effect and information source. Here, the results showed a multimedia effect only for picture questions.

The results of the first experiment seem to support the account of computational offloading best, as the multimedia effect differed between task types and the multimedia effect was larger in procedural tasks than in conceptual tasks, for free recall and recall verification accuracy. The results also seem to support CTML as the multimedia effect was equally large in causal and conceptual tasks.

However, the results from Experiment 1 were not always as expected by either CTML or the account of computational offloading. One unexpected result was that the multimedia effect was larger in procedural tasks than in causal tasks. An explanation for this result could be that there were overall differences between procedural versus causal tasks in the challenges imposed onto learners. It could be argued that procedural tasks were more complex than causal tasks, because procedural tasks required a 3D spatial mental representation. Due to these differences in complexity, the instructional support provided by the pictures — as opposed to the more error prone process of imagery — may have had a larger impact for the more complex procedural tasks. This may have been the case particularly because the procedural pictures provided depth cues (e.g., a rope goes behind a branch), which supported learners to construct the required 3D spatial mental representation of the procedure.

Another unexpected result was that for integration verification accuracy questions the pattern of the size of the multimedia effect was exactly opposite. Here, the multimedia effect was largest in conceptual tasks and smallest in procedural tasks. A possible explanation for the small and non-existing multimedia effect in causal and procedural tasks respectively, is that for the questions in these two task types, it was easy enough to guess the correct answer compared to the corresponding questions for conceptual tasks. This interpretation receives some support when looking at the high means for this measure of students in the text-only conditions (chance level: .50; conceptual: M = .66; causal: M = .79; procedural: M = .71). Even though students in the text-only condition never saw the picture, which was required to answer these items, for causal and procedural tasks they nevertheless scored high, suggesting that no pictorial support was required. This was not the case for performance in the text-only condition for conceptual tasks, where students' performance was only slightly above chance level, suggesting that here there was opportunity for improvement. Accordingly, pictures aided integration verification performance for conceptual tasks, however less so for causal and procedural tasks.

A final unexpected result was that a multimedia effect was found for transfer questions only when considering picture questions. This result can be seen as surprising, because Mayer (2009b) repeatedly found a multimedia effect for transfer questions. A possible explanation for this discrepancy is that in Mayer's and other experiments, information source was not taken into account. It is possible that the multimedia effect in these studies is also found only for questions that heavily rely on information provided only in the picture. However, as typically no distinction is made between information sources, this cannot be examined. Regardless of whether this is indeed the case, the results from this experiment show that incorrect conclusions can be made concerning the beneficial effect of pictures, when information source is not taken into account. Therefore, similar to Levie and Lentz (1982), I suggest that future studies, if possible, start taking this distinction into account.

Taken together, from a theoretical point of view, the question whether the multimedia effect can be explained best by CTML or the account of computational offloading cannot be fully answered yet. In Experiment 2, this question is again addressed by focusing on whether information processing in working memory differs between task types.

3 Experiment 2

To further investigate the relative contribution of CTML and the account of computation offloading for explaining the multimedia effect, working memory involvement during learning from text and pictures was gauged using the dual task approach to assess whether differences in information processing during learning account for the differences between task types. Accordingly, the purpose of the second experiment was to investigate working memory involvement when studying conceptual, causal, and procedural, using the same learning materials and outcome measures as in Experiment 1. Here, CTML, which focuses on dual coding, would predict that the central executive, the phonological loop, and the visuo-spatial sketchpad are involved when learning with multimedia, and that there are no differences between conceptual, causal, and procedural tasks in working memory involvement. In contrast, the account of computational offloading, which focuses on facilitated information processing, would predict that the involvement of the visuo-spatial sketchpad is higher when visuo-spatial information is more important for understanding the learning task. Therefore, it would predict that the involvement of the visuo-spatial sketchpad is higher in causal and procedural tasks than in conceptual tasks. No differences between conceptual, causal, and procedural tasks would be predicted concerning the involvement of the central executive and phonological loop.

3.1 Method

3.1.1 Participants and design

121 participants (95 female and 26 male; $M = 24.23$ years, $SD = 3.70$) were randomly assigned to one of five conditions. A 5x3x3 design was used, with dual task (i.e., with vs. without) as between-subject variable, and task type (i.e., conceptual vs. causal vs. procedural) and information source (i.e., text only vs. text-pictures vs. picture questions) as within-subject variable. Depending on condition, participants performed 1) no dual task (i.e., control condition; $n = 25$), 2) a dual task to load the phonological loop (i.e., PL condition; $n = 25$), 3) a dual task to load the visuo-spatial dual task (i.e., VSSP condition; $n = 24$), 4) a dual task to load the central executive with a verbal component (i.e., CE-PL condition; $n = 22$), or 5) a dual task to load the central executive with a visuo-spatial component (i.e., CE-VSSP condition; $n = 25$). As in Experiment 1, all participants studied materials of the three task

types. Also, for each task, the same sources of information were defined. Participants received either payment or course credit.

3.1.2 Materials and measures

In Experiment 2, the same materials were used as in Experiment 1 with a few exceptions.

3.1.2.1 Multimedia learning materials

The six tasks (two from each task type) that seemed most beneficial to learning in Experiment 1 and in which integration verification questions were least easy to guess were again used in Experiment 2. Only six tasks were used to reduce the required number of participants. The selected conceptual tasks concerned the relationships between inhabitants of a village (see Figure 2) and an animals' food hierarchy. The causal tasks concerned a machine to notify that eggs should be collected (see Figure 3) and a machine to warn for danger. The procedural tasks concerned how to bend a tree to obtain drinkable juice (see Figure 4) and about how to sew a headscarf. For the selected tasks, the total number of propositions (i.e., in text and in pictures) in conceptual tasks was on average 128, in causal tasks 148, and in procedural tasks 127. For all tasks, the number of propositions only mentioned in the text ranged from 68 to 75, only in the picture from 28 to 66, and in both text and picture (i.e., overlap) from 10 to 18.

3.1.2.2 Post-tests

The same post-test items as in Experiment 1 were used (see Section 2.1.2.2). In addition, a free transfer test was conducted. In the free transfer test, participants answered three open transfer questions. A scoring system was created that contained the information constituting a correct answer. Based on this scoring system, each answer was allotted two points when the answer was correct, one point was allotted when the answer was partly correct, and zero points were allotted when the answer was incorrect. For example, the correct answer to the question "Who has to join so that all men are playing headball together?" was Yellow Rectangle. Two points were allotted when the answer was "Yellow Rectangle". One point was allotted when "Yellow Rectangle", but also one or more other persons were mentioned. Also, one point was allotted when only part of the name was mentioned and thereby causing ambiguity. Zero points were allotted when "Yellow Rectangle" was not mentioned. To assess the inter-rater reliability, two raters coded all answers. In cases where the raters disagreed, the raters discussed the answer and came to a joint decision.

3.1.2.3 Dual tasks

To load the three working memory components, four dual tasks were used. To load the phonological loop (PL), the number repetition task was used, in which participants spoke the number 1 to 4 in the same order, one number per second, during learning (e.g., Baddeley, Eldridge, & Lewis, 1981). To load the visuo-spatial sketchpad (VSSP), the foot tapping task was used, in which participants tapped four foot pedals on the floor with their right foot in clockwise order and one pedal per second during learning (e.g., Miyake et al., 2004). To load the central executive (CE), the random number repetition (based on Brunyé et al., 2006; Baddeley et al., 1981) and random foot tapping tasks (Brunyé et al., 2006) were used, in which participants spoke the numbers 1 to 4 or tapped the four foot pedals in random order, respectively. The pedals for the foot tapping tasks were arranged in a rectangle, which was 27 cm wide and 32 cm long. The size of each pedal was 6 cm wide and 9 cm long.

3.1.3 Procedure

The procedure for Experiment 2 was similar to the procedure of Experiment 1. One difference was that participants who had to perform a dual task were instructed to perform this dual task during studying the instructional material. They were informed that if they would stop speaking numbers or stop pressing the pedals for five seconds, a beep would remind them to keep on performing the dual task. After these instructions, they practiced their dual task for two minutes, while hearing a metronome that indicated the length of a second. Only participants in the dual task conditions performed the dual task while studying all six learning tasks. Another difference was that participants performed the free transfer test after the other learning outcome measures. They were allowed a maximum of 5 minutes to answer all three questions. The duration of one session was between 2.5 and 3 hours.

3.1.4 Data analysis

To be able to investigate the role of working memory in conceptual, causal, and procedural tasks, for different information sources when considering different types of learning outcomes, five mixed ANOVAs were conducted. The dependent variables for these analyses were free recall, recall verification accuracy, free transfer, transfer verification accuracy, and integration verification accuracy. The between-subjects factor was dual task. The dual task conditions were the control, speaking, foot tapping, random speaking, and random foot tapping condition tasks. In the speaking condition, the phonological loop was loaded by the dual task (i.e., PL condition), in the foot tapping condition the visuo-spatial sketchpad (i.e., VSSP condition), in the random speaking condition the verbal part of the central executive (i.e., CE-PL condition), and in the random foot tapping task the visuo-spatial part of the central executive (i.e., CE-VSSP condition). As within-subjects factor, task type (i.e., conceptual, causal, procedural) was included. For free recall, recall verification accuracy, and transfer verification accuracy, information source (i.e., information given in text only, text and picture, or picture only) was included as a second within-subjects factor. As for Experiment 1, the main effects of task type and of information source, and the interaction between task type and information source are not reported.

Furthermore, a mixed ANOVA was conducted with study time as dependent variable, task type as within-subjects factor, and dual task as between subjects-factor, to make sure that possible differences between conditions were not due to differences in study times.

All statistical assumptions for the reported tests were met. All effects are reported as significant at $p < .05$. Partial eta-squared was used as effect size. For significant overall effects, Bonferroni post-hoc tests were conducted.

3.2 Results

Descriptive statistics are provided in Table 4.

3.2.1 Free recall

The results of the mixed ANOVA with free recall as dependent variable showed a main effect of dual task ($F(4, 116) = 5.10$, $p < .01$, $\eta_p^2 = .15$). Post-hoc comparisons showed that learning outcomes in the control condition (*M* = .58, *SD* = .11) where higher than in the CE-PL condition (*M* = .45, *SD* = .15; $p < .01$), but were the same as in the PL condition (*M* = .55, *SD* = .11; $p = 1.00$), the VSSP condition (*M* = .57, *SD* = .10; $p = 1.00$), and the CE-VSSP condition (*M* = .59, *SD* = .12; $p = 1.00$). This implies that the dual task only interfered with information processing during learning when the dual task loading the central executive had a verbal component. Furthermore, learning outcomes in the CE-PL condition were lower than in all other conditions ($p < .05$).

The results did not provide evidence for a significant 2-way interaction between task type and dual task ($F(8, 230) = 1.56$, $p = .14$, $\eta_p^2 = .05$), a 2-way interaction between information source and dual task ($F(8, 230) = 1.88$, $p = .06$, $\eta_p^2 = .06$), nor a 3-way interaction between information source, task type, and dual task ($F(16, 346) = 1.22$, $p = .25$, $\eta_p^2 = .04$), implying that the differences between dual task conditions were the same in each task type for each information source.

3.2.2 Recall verification accuracy

The results of the mixed ANOVA with recall verification accuracy as dependent variable showed a main effect of dual task ($F(4, 116) = 8.27$, $p < .001$, $\eta_p^2 = .22$). Post-hoc comparisons showed that learning outcomes in the control condition ($M = .81$, $SD = .06$) were higher than in the CE-PL condition ($p < .001$; $M = .72$, $SD = .07$), but were the same as in the PL condition ($M = .79$, $SD = .06$; $p = 1.00$), the VSSP condition ($M = .79$, $SD = .07$; $p = 1.00$), and the CE-VSSP condition ($p = 1.00$; $M = .82$, $SD = .06$). This implies that — again — the dual task only interfered with information processing during learning when the dual task loading the central executive had a verbal component. Furthermore, learning outcomes in the CE-PL condition were lower than in all other conditions ($p < .01$).

The results also demonstrated a significant 2-way interaction between task type and dual task ($F(8, 230) = 2.12$, $p = .03$, $\eta_p^2 = .07$), implying that the difference between dual task conditions differed between task types. Post-hoc comparisons provided evidence that for conceptual tasks, learning outcomes in the CE-PL condition were lower than in the remaining conditions (control condition: $M = .79$, $SD = .08$; PL condition: $M = .79$, $SD = .10$; VSSP condition: $M = .81$, $SD = .09$; CE-PL condition: $M = .70$, $SD = .13$; CE-VSSP condition: $M = .82$, $SD = .10$; $p = .03$). In causal task, learning outcomes in the control condition were higher than in the CE-PL condition (control condition: $M = .86$, $SD = .08$; CE-PL condition: $M = .71$, $SD = .09$; $p < .001$), and higher than in the VSSP condition ($M = .77$, $SD = .10$; $p = .02$); the control condition did not differ from the PL condition ($M = .81$, $SD = .10$; $p = .59$), nor from the CE-VSSP condition ($M = .83$, $SD = .09$; $p = 1.00$). Also, learning outcomes in the CE-PL condition were lower than in all other conditions (p ranging from < .001 to .01) except from the VSSP condition ($p = .47$). In procedural tasks, learning outcomes were the same in all conditions ($p = 1.00$). Also, there were no differences between the CE-PL condition and the other conditions. This implies that in conceptual and causal tasks, the dual task interfered with information processing during learning when the dual task loading the central executive had a verbal component. This also means that only in causal tasks the learning task interfered with information processing in the visuo-spatial sketchpad. There were no further differences between dual task conditions in any of the tasks types other than the ones described here.

The results did not provide evidence for a significant 2-way interaction between information source and dual task ($F < 1$), implying that the differences between dual task conditions were the same for all information sources, nor for a 3-way interaction between information source, task type, and dual task ($F < 1$), implying that the interaction between dual task and task type was the same for all information sources.

3.2.3 Free transfer

The results of the mixed ANOVA with free transfer as dependent variable did not provide evidence for a main effect of dual task ($F(4, 116) = 2.34$, $p = .06$, $\eta_p^2 = .08$), nor for a 2-way interaction between task type and dual task ($F < 1$).

3.2.4 Transfer verification accuracy

The results of the mixed ANOVA with free transfer verification accuracy as dependent variable showed a main effect of dual task ($F(4, 116) = 3.18$, $p = .02$, $\eta_p^2 = .10$), implying that there were differences between the dual task conditions. However, post-hoc comparisons demonstrated that there were no differences between the control condition and any of the experimental conditions (control condition: $M = .70$, $SD = .08$; PL condition: $M = .70$, $SD = .09$; VSSP condition: $M = .73$, $SD = .07$; CE-PL condition: $M = .66$, $SD = .09$; CE-VSSP condition: $M = .74$, $SD = .07$; $p = 1.00$). This means that processing the dual task did not interfere with information processing in any of the learning tasks. Furthermore, learning outcomes in the CE-PL condition were lower than in the VSSP condition ($p = .04$) and the CE-VSSP condition ($p = .02$).

The results did not provide evidence for a significant 2-way interaction between task type and dual task ($F(8, 230) = 1.42$, $p = .19$, $\eta_p^2 = .05$), nor for a 2-way interaction between information source and dual task ($F(8, 230) = 1.02$, $p = .30$, $\eta_p^2 = .04$), nor for a 3-way interaction between information source, task type, and dual task ($F < 1$), implying that the differences between dual task conditions were the same in each task type for each information source.

3.2.5 Integration verification accuracy

The results of the mixed ANOVA with integration verification accuracy as dependent variable showed no main effect of dual task ($F(4, 116) = 1.38$, $p = .25$, $\eta_p^2 = .05$), implying that there were no differences between the dual task conditions.

The results showed a significant 2-way interaction between task type and dual task ($F(8, 230) = 2.12$, $p = .04$, $\eta_p^2 = .07$). Post-hoc comparisons demonstrated that in conceptual tasks (control condition: *M* = .77, *SD* = .17; PL condition: *M* = .72, *SD* = .20; VSSP condition: *M* = .79, *SD* = .14; CE-PL condition: *M* = .73, *SD* = .12; CE-VSSP condition: *M* = .82, *SD* = .17; *p* = 1.00) and procedural tasks (control condition: *M* = .73, *SD* = .21; PL condition: *M* = .77, *SD* = .13; VSSP condition: *M* = .78, *SD* = .16; CE-PL condition: *M* = .80, *SD* = .14; CE-VSSP condition: *M* = .74, *SD* = .20; *p* = 1.00), there were no differences between any of the dual task conditions. For causal tasks, the control condition (*M* = .73, *SD* = .21) also did not differ from any of the experimental conditions (PL condition: *M* = .77, *SD* = .13, *p* = 1.00; VSSP condition: *M* = .78, *SD* = .16, *p* = 1.00; CE-PL condition: *M* = .80, *SD* = .14, *p* = .06; CE-VSSP condition: *M* = .74, *SD* = .20, *p* = 1.00), however, learning outcomes in the CE-PL condition were higher than in the CE-VSSP condition ($p < .01$). The comparisons between the control condition and experimental conditions imply that performing the dual task did not interfere with information processing during learning in any of the learning tasks.

Table 4

Means and Standard Deviations for Study Times, Free Recall, Recall Verification Accuracy, Free Transfer, Transfer Verification Accuracy, and Integration Verification Accuracy as a Function of Picture-presentation condition, Information Source, and Task Type

	Picture-presentation condition	Control condition			PL condition			VSSP condition			CE-PL condition			CE-VSSP condition		
	Information source	Text	Text-Picture	Picture	Text	Text-Picture	Picture	Text	Text-Picture	Picture	Text	Text-Picture	Picture	Text	Text-Picture	Picture
Dependent variable	Task type	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*	*M* *(SD)*
Free recall (proportion recalled)	Conceptual	.53 (.14)	.81 (.12)	.58 (.16)	.53 (.12)	.82 (.11)	.53 (.14)	.56 (.15)	.84 (.10)	.57 (.13)	.41 (.17)	.68 (.16)	.42 (.18)	.57 (.17)	.84 (.14)	.58 (.19)
	Causal	.39 (.11)	.86 (.14)	.48 (.13)	.37 (.09)	.82 (.15)	.43 (.16)	.38 (.11)	.83 (.16)	.44 (.14)	.29 (.13)	.69 (.21)	.34 (.15)	.40 (.08)	.87 (.09)	.51 (.11)
	Procedural	.37 (.09)	.78 (.16)	.46 (.16)	.34 (.10)	.73 (.17)	.40 (.13)	.38 (.11)	.75 (.12)	.40 (.11)	.29 (.11)	.61 (.23)	.33 (.15)	.39 (.12)	.75 (.17)	.41 (.15)
Recall verification accuracy (proportion correct)	Conceptual	.86 (.11)	.90 (.12)	.61 (.16)	.85 (.15)	.89 (.09)	.63 (.19)	.89 (.11)	.94 (.08)	.59 (.25)	.76 (.19)	.85 (.17)	.50 (.19)	.89 (.12)	.92 (.12)	.65 (.23)
	Causal	.88 (.14)	.93 (.10)	.77 (.14)	.81 (.16)	.91 (.11)	.71 (.18)	.78 (.14)	.85 (.18)	.69 (.18)	.67 (.18)	.84 (.11)	.64 (.16)	.83 (.17)	.91 (.13)	.75 (.19)
	Procedural	.93 (.10)	.72 (.22)	.72 (.17)	.93 (.19)	.73 (.18)	.70 (.16)	.93 (.19)	.69 (.16)	.73 (.15)	.86 (.15)	.63 (.20)	.78 (.15)	.93 (.12)	.71 (.17)	.77 (.16)
Free transfer (proportion recalled)	Conceptual		.58 (.24)			.47 (.20)			.47 (.23)			.44 (.21)			.55 (.25)	
	Causal		.68 (.17)			.68 (.15)			.61 (.24)			.56 (.25)			.67 (.18)	
	Procedural		.67 (.21)			.65 (.19)			.61 (.18)			.56 (.20)			.65 (.20)	

Transfer verification	Conceptual	.75	.56	.53	.70	.63	.55	.78	.63	.56	.66	.64	.55	.73	.61	.63
accuracy (proportion		(.18)	(.18)	(.21)	(.14)	(.14)	(.20)	(.17)	(.16)	(.19)	(.23)	(.25)	(.27)	(.12)	(.18)	(.19)
correct)	Causal	.65	.75	.69	.71	.81	.69	.72	.74	.72	.66	.71	.63	.78	.75	.69
		(.18)	(.17)	(.21)	(.15)	(.17)	(.24)	(.14)	(.16)	(.15)	(.)22	(.16)	(.23)	(.16)	(.20)	(.16)
	Procedural	.73	.78	.87	.69	.73	..69	.78	.81	.83	.63	.70	.77	.79	.77	.86
		(.15)	(.19)	(.14)	(.20)	(.21)	(.15)	(.20)	(.17)	(.16)	(.21)	(.21)	(.15)	(.18)	(.14)	(.12)
Integration verification	Conceptual		.80			.72			.79			.73			.82	
accuracy (proportion			(.17)			(.20)			(.14)			(.12)			(.17)	
correct)	Causal		.84			.81			.81			.71			.88	
			(.16)			(.13)			(.15)			(.21)			(.13)	
	Procedural		.73			.77			.78			.80			.74	
			(.21)			(.13)			(.16)			(.14)			(.20)	
Study times (min)	Conceptual		3.55			3.51			3.28			3.60			3.89	
			(1.27)			(1.12)			(1.03)			(1.66)			(1.46)	
	Causal		2.78			3.01			2.79			3.06			3.07	
			(0.97)			(0.87)			(0.99)			(1.00)			(1.08)	
	Procedural		3.29			3.06			2.79			2.97			2.98	
			(1.30)			(0.98)			(0.96)			(0.99)			(1.26)	

3.2.6 Study times

The results of the ANOVA with study times as dependent variable did not show a main effect of dual task ($F < 1$). Also, there was no 2-way interaction between task type and dual task ($F(8, 230) = 1.66$, $p = .11$, $\eta_p^2 = .06$).

3.3 Discussion

The aim of the second experiment was to investigate whether working memory involvement when learning with text and pictures can account for the differences concerning the beneficial effect of pictures between task types. Therefore, working memory involvement, when learning with text and pictures during studying conceptual, causal, and procedural tasks, was assessed. Different learning outcome measures (i.e., free recall, recall verification accuracy, free transfer, transfer verification accuracy, and integration verification accuracy) and three information sources (i.e., information given in text only, in text and picture, or in picture only) were used. Participants learned without dual tasks, or with one of four dual tasks loading one of the working memory subsystems. CTML would predict that the central executive, the phonological loop, and the visuo-spatial sketchpad are involved when learning with multimedia, and that there are no differences between conceptual, causal, and procedural tasks in working memory involvement. In contrast, the account of computational offloading would predict that the involvement of the visuo-spatial sketchpad is higher when visuo-spatial information is more important for understanding the learning task. Therefore, it would predict that the involvement of the visuo-spatial sketchpad is higher in causal and procedural tasks than in conceptual tasks. No differences between conceptual, causal, and procedural tasks would be predicted concerning the involvement of the central executive and phonological loop.

The results of the second experiment do not seem to support the account of computational offloading and seem to support CTML best at first sight, as the results showed that learning outcomes from participants in the control condition did not differ from the learning outcomes from participants who performed a dual task regardless of information source and task type. There were two exceptions to this general pattern: for free recall, learning outcomes in the control condition were higher than in the condition in which the dual task loaded the central executive with a verbal component; for recall verification accuracy, this result was also found for conceptual and causal tasks and learning outcomes in the control condition were higher than in the condition in which the dual task loaded the visuo-spatial sketchpad. These results do not seem to explain the beneficial effect of pictures from Experiment 1, as in Experiment 1 free recall and recall verification accuracy showed a larger multimedia effect in procedural tasks than in conceptual and causal tasks.

Even though the results seem to be more in line with CTML, it is unclear whether the lack of effects can be interpreted in favour of CTML, as there were several unexpected patterns of results. Firstly, performing a dual task that loaded the central executive having a verbal component interfered with information processing during learning, whereas performing a dual task that loaded the central executive having a visuo-spatial component did not. If the central executive was indeed involved during learning, both dual tasks loading the central executive should have evoked the same pattern of results. Alternatively, if not the central executive but instead the verbal component of the dual task was responsible for eliciting interference with the learning task, also the dual task loading the phonological loop should have interfered with information processing learning, which was not the case. Secondly, performing a dual task loading the phonological loop, the visuo-spatial sketchpad, and the central executive with a visuo-spatial component did not seem to interfere with information processing during learning. As previous research has shown that these working memory subsystems are involved during learning, a possible explanation for this discrepancy could be that in this study only the central executive dual task with a verbal component was difficult enough to cause overload in the central executive. In contrast, the other dual tasks might have been easy enough, so that they could be performed relatively effortless, and therefore did not cause overload in the respective working memory subsystem. This postulation is supported by the result that not only learning outcomes but also study times were similar between conditions.

A possible limitation of Experiment 1 and 2 is that the difficulty of both post-test questions might have influenced the pattern of result. I tried to reduce the varying difficulty between questions and learning tasks, by using at least three questions for every dependent variable and two or three tasks per task type. Unfortunately, nothing can be said concerning the absolute difficulty of a task as performance on any test item always results from interplay between the difficulty of the task and of the item itself. When comparing different materials, each having own post-test items, it is impossible to rule out this interplay completely.

Another possible limitation of Experiment 1 and 2 is that learning outcomes in all task types were gauged using recall of the structure and interrelations between objects and transfer of information to new situations. These learning outcome measures are suitable for conceptual and causal tasks. However, for procedural tasks, it would be more appropriate to test how well the procedure is performed. Therefore, in future research, it is important to measure the beneficial effect of pictures in procedural tasks by gauging performance accuracy as learning outcome measure.

Taken together, working memory involvement does not seem to differ between task types as would be expected based on the size of the multimedia effect in different task types. However, dual task performance had unexpectedly little effect on learning outcomes in general. Therefore, a more direct test whether pictures are computational offloading should provide more insight into whether the account of computational offloading can predict the beneficial effect of pictures.

4 Experiment 3[3]

As the first two experiments have not been able to provide a clear pattern of results that can be used to contrast CTML and the account of computational offloading, the third experiment specifically aims at testing whether pictures are computational offloading. Here, the potential limitations from the first two experiments concerning features of the tasks and the dependent variables are taken into account. Accordingly, the purpose of the second experiment was to investigate the involvement of the visuo-spatial sketchpad when learning with text only or with text and pictures in procedural tasks that are assumed to trigger imagery (i.e., first aid procedures), by using the same post-test for all learning tasks and assessing performance accuracy. CTML would predict that the visuo-spatial sketchpad is more involved when learning with text and pictures than when learning with text only, as the visuo-spatial sketchpad is involved in picture processing. In contrast, the account of computational offloading would predict that the visuo-spatial sketchpad is more involved when learning with text only than when learning with text and pictures, as imagery requires more mental resources than visuo-spatial perception.

4.1 Method

4.1.1 Participants and design

Eighty-seven students from a Dutch university studied four first aid procedures. Five participants were excluded, as dual task performance showed that they had not followed the instructions properly. Of the remaining 82 participants (66 female and 16 male; *M* = 22.80 years, *SD* = 3.05), 78% had no prior experience with first aid instructions. The experiment had a 2x2x2 mixed design, with presentation format (i.e., text-only vs. multimedia) and dual task (i.e., with vs. without) as between-subject variables. We included time of testing (immediate vs. delayed) as a within-subject variable to see whether effects were the same for both immediate and delayed testing. Depending on presentation format and dual task condition, participants received 1) a text without dual task (*n* = 21), 2) a text with a dual task

[3] This article has been submitted as: van Genuchten, E., van Hooijdonk, C., Schüler, A., & Scheiter, K. (submitted). The beneficial effect of pictures and the role of working memory when "learning how" with multimedia learning material. *Applied Cognitive Psychology*.

(*n* = 20), 3) a multimedia instruction without dual task (*n* = 21), or 4) a multimedia instruction with a dual task (*n* = 20). Learning outcomes were measured immediately after learning and again after one week. Participants were randomly assigned to one of the four experimental conditions and received either payment or course credit for their participation.

4.1.2 Materials and measures

4.1.2.1 Learning tasks

The procedural tasks were four first aid instructions obtained from the Orange Cross manual (Henny, 2006). First aid tasks were used as these tasks are procedural tasks conveying actions by describing/depicting the actor's body parts required to perform the actions (see Figure 5 for an example). The tasks described a) how to fold a sling in supporting a broken arm across a victim's chest, b) how to roll a victim from the recovery position onto their back, c) how to apply an easy-application bandage, and d) how to move an unconscious victim from areas of danger. The tasks contained 4, 5, 10, 12 steps and 52, 51, 114, 107 words, respectively. In the multimedia instructions, two or three pictures accompanied the text. In all tasks, the steps, which are required to perform the procedure correctly, and how these steps should be executed, were described in the text. Pictures were photographs in which both the object (i.e., the victim) and the person performing actions were depicted. In the text, action-related words were used (e.g., "sit down in squat position with your feet at each side of the victim and as close to the victim as possible"). Studying the text only allowed correct performance of the procedure. The text was presented to the left of the pictures when learning with multimedia, whereas the text was presented in the middle of the screen when learning with text only. The first aid tasks were presented in random order using E-prime v.1.2 (Schneider, Eschman, & Zuccolotto, 2002) on a PC computer with 22 inch monitor.

4.1.2.2 Dual task

The dual task used to load the visuo-spatial sketchpad during learning was, as in Experiment 2, the foot tapping task (see Section 3.1.2.3; e.g., Miyake et al., 2004).

4.1.2.3 Post-tests

To test how much knowledge participants had obtained, two retention tests were used. In the first test, which focused on procedural-motor/implicit knowledge, participants executed the first aid tasks that they had studied using a first aid dummy. This dummy was sitting in a chair for the two bandaging tasks and lying on the floor for the other two tasks. Bandage materials were provided when required for performing the task. Participants did not receive any feedback on their performance. Performance accuracy was measured by the proportion of steps that were performed both correctly and in the correct order. Participants could score either 0 or 1 per step. The proportion of correctly performed steps was calculated, resulting in one score between 0 and 1. To assess the inter-rater reliability for performance accuracy, two raters coded 20% of all videos. Cohen's kappa was .71. The remaining 80% of the data were scored by a single rater only.

In the second test, which focused on declarative/explicit knowledge, participants saw a picture from the studied task and verbally described the steps that either preceded or followed the depicted step. Description accuracy was measured by the proportion of steps that were described both correctly and in the correct order. Participants could score either 0 or 1 per step. The proportion of correctly described steps was calculated, resulting in one score between 0 and 1. To assess the inter-rater reliability for description accuracy, two raters coded 20% of all texts. Cohen's kappa was .67. The remaining 80% of the data were scored by a single rater only.

4.1.3 Procedure

Before the experiment started, participants gave their informed consent. Then, they familiarised themselves with performing a simple first aid task that was unrelated to understanding of the experimental learning tasks on a first aid dummy. Participants received a written task on paper, which described in three steps how to tilt a patient onto the side and back. Participants in the text-only conditions received this training task without a picture, whereas participants in the multimedia conditions received this task with pictures.

Subsequently, all participants answered a demographic questionnaire (i.e., age, sex, education, prior knowledge concerning first aid procedures). Participants who had to perform a dual task were instructed to press the foot pedals in clockwise order and one pedal per second whilst studying the learning material. They were informed that if they would stop pressing the pedals for five seconds, a beep would remind them to continue pressing the pedals. After these instructions, they practiced the foot tapping task for one minute, while hearing a metronome that indicated the length of a second. Then, all participants were directed to study the four first aid tasks and were informed that after studying these tasks, they would be tested on their acquired knowledge without referring back to the learning material. No time limit for studying was set so that participants could continue with the next task when they felt confident that they had understood the task and remembered all information. Participants in the dual task conditions performed the dual task whilst studying all four learning tasks. After participants in all conditions finished studying a first aid task, they continued to the next task by pressing a key on a keyboard.

After learning the first aid tasks, participants executed these tasks using the first aid dummy. During their performance, participants' actions were recorded with a video camera from two angles. Also, a photograph was taken after the easy-application bandage had been applied around the first aid dummy's arm. After executing a first aid task, participants received a picture from the learning material and described the preceding or subsequent steps. This procedure was repeated for each task, in the same order as they had been studied. One week later, students performed the same learning outcome tests in the same order in the same room. No time limits were set for executing the post-tests. The first session took between 45 and 60 minutes and the second session about 30 minutes. Each participant was tested individually.

4.1.4 Data analysis

To investigate whether pictures are computational offloading, two mixed ANOVAs were performed. The dependent variables for these analyses were performance accuracy and description accuracy. The between-subject factors were presentation format (text-only vs. multimedia) and dual task (with vs. without). As within-subject factor, time of testing (immediate vs. delayed) was included. The learning outcome measures for immediate and delayed testing for each task were standardised to reduce the effect of the different number of steps between tasks. The partial eta-squared effect size is reported to reflect the amount of variance that is explained by the model after other non-error sources of variance have been partialled out (Cohen, 1973). For partial eta-squared, .01, .06, and .14 correspond to small, medium, and large effect sizes, respectively (Cohen, 1988). To follow up on significant interactions ($p < .05$), Bonferroni adjusted pairwise comparisons were conducted.

4.2 Results

Descriptive statistics are provided in Table 5.

4.2.1 Main effects

The results showed a main effect of presentation format for both performance accuracy and description accuracy, implying that participants in the multimedia condition performed and described the procedures correctly more often than participants in the text-only condition (performance accuracy: $F(1, 78) = 34.14$, $p < .001$, $\eta_p^2 = .30$; text-only condition: $M = -.39$, $SE = .09$; multimedia condition, $M = .38$, $SE = .09$; description accuracy: $F(1, 78)$ 4.97, $p = .03$, $\eta_p^2 = .06$; text-only condition, $M = -.19$, $SE = .11$; multimedia condition: $M = .17$, $SE = .11$). These results imply that a multimedia effect was found for both dependent variables. However, the main effects of presentation format were qualified by interactions described below.

The results also showed a main effect of dual task for both performance accuracy and description accuracy, implying that participants who did not perform the dual task performed and described the procedures correctly more often than participants who did perform the dual task (performance accuracy: $F(1, 78) = 22.09$, $p < .001$, $\eta_p^2 = .22$; without dual task: $M = .30$, $SE = .09$; with dual task, $M = -.32$, $SE = .09$; description accuracy: $F(1, 78) = 14.47$, $p < .001$, $\eta_p^2 = .16$; without dual task: $M = .30$, $SE = .11$; with dual task, $M = -.31$, $SE = .11$). However, the main effect of dual task was qualified by interactions described below.

The results did not show a main effect of time of testing (performance accuracy: $F < 1$; description accuracy: $F < 1$), showing that participants performed and described the procedures equally correctly for both immediate and delayed testing.

4.2.2 Two-way interactions

The results showed a significant 2-way interaction between presentation format and dual task for performance accuracy ($F(1, 78) = 9.27$, $p < .01$, $\eta_p^2 = .11$; see Figure 9), however not for description accuracy ($F(1, 78) = 3.26$, $p = .08$, $\eta_p^2 = .04$). Bonferroni adjusted pairwise comparisons showed that performing a dual task interfered with information processing during learning with text only ($p < .001$, $\eta_p^2 = .28$; text-only without dual task: $M = .12$, $SE = .13$; text-only with dual task: $M = -.90$, $SE = .13$), but did not interfere with information processing during learning with multimedia (p = .25, η_p^2 = .02; multimedia without dual task: M = .49, SE = .13; multimedia with dual task: M = .27, SE = .13).

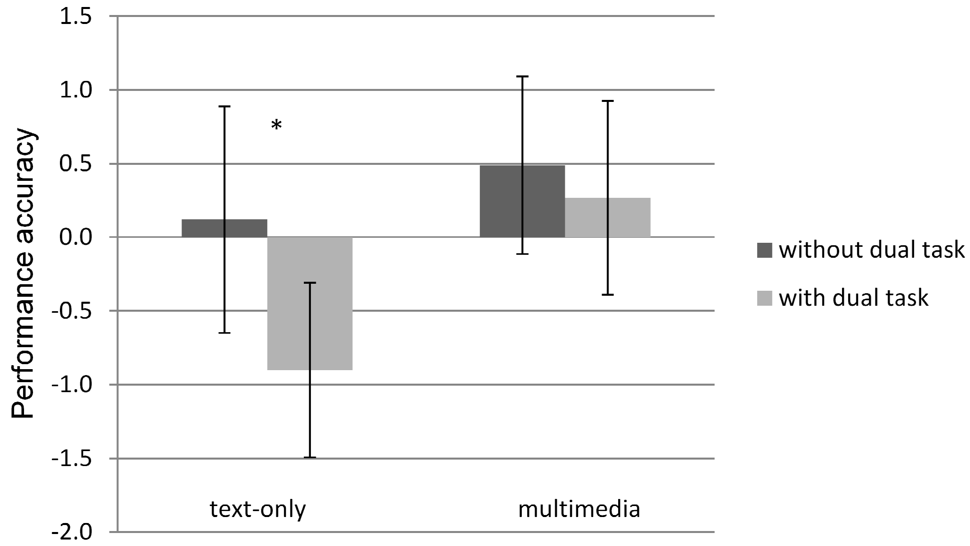

Figure 9. Means and standard deviations for the text-only and multimedia conditions for performance accuracy as a function of dual task.

The results also showed a significant 2-way interaction between presentation format and time of testing for performance accuracy (F(1, 78) = 6.96, p = .01, η_p^2 = .08; see Figure 10), but not for description accuracy (F < 1). Bonferroni adjusted pairwise comparisons for performance accuracy showed that the multimedia effect was smaller for delayed testing then for immediate testing (immediate: p < .001, η_p^2 = .38; text-only: M = -.48, SE = .10; multimedia: M = .46, SE = .10; delayed: p < .001, η_p^2 = .16; text-only: M = -.31, SE = .11; multimedia: M = .29, SE = .11).

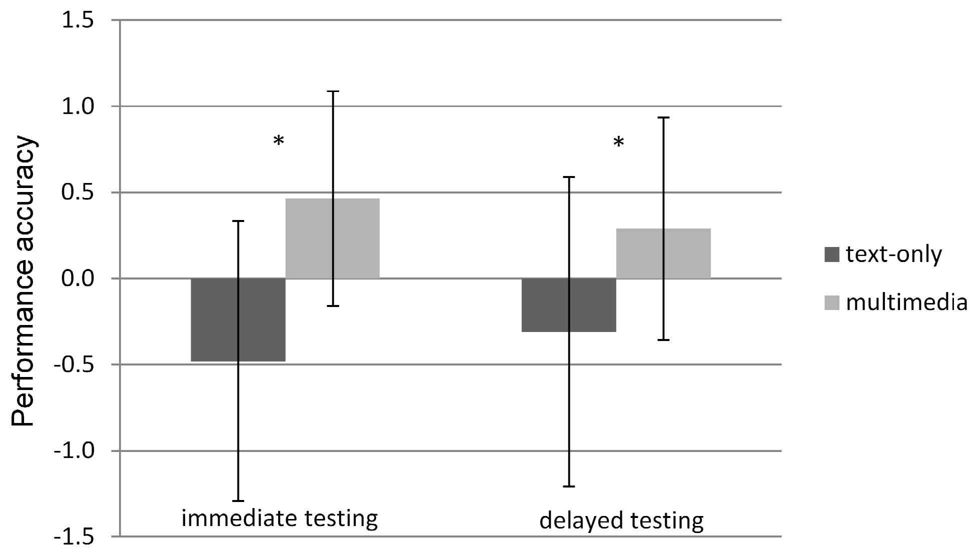

Figure 10. Means and standard deviations for the text-only and multimedia conditions for performance accuracy as a function of time of testing.

The results did not show a significant 2-way interaction between dual task and time of testing (performance accuracy: $F < 1$; description accuracy: $F < 1$).

4.2.3 Three-way interactions

The results showed a significant 3-way interaction between presentation format, dual task, and time of testing for description accuracy, but not for performance accuracy (performance accuracy: $F < 1$; description accuracy: $F(1, 78) = 6.47$, $p = .01$, $\eta_p^2 = .08$; see Figure 11). Bonferroni adjusted pairwise comparisons for description accuracy showed that for immediate testing performing a dual task affected learning outcomes only when learning with text only (text-only: $p < .001$, $\eta_p^2 = .20$; without dual task: $M = .37$, $SE = .17$; with dual task: $M = -.68$, $SE = .17$; multimedia: $p = .57$, $\eta_p^2 < .01$; without dual task: $M = .21$, $SE = .17$; with dual task: $M = .08$, $SE = .17$). However, for delayed testing, performing a dual task affected learning outcomes both when learning with text-only and when learning with multimedia (text-only: $p < .01$, $\eta_p^2 = .10$; without dual task: $M = .16$, $SE = .18$; with dual task: $M = -.59$, $SE = .18$; multimedia: $p < .05$, $\eta_p^2 = .05$; without dual task: $M = .45$, $SE = .18$; with dual task: $M = -.05$, $SE = .18$).

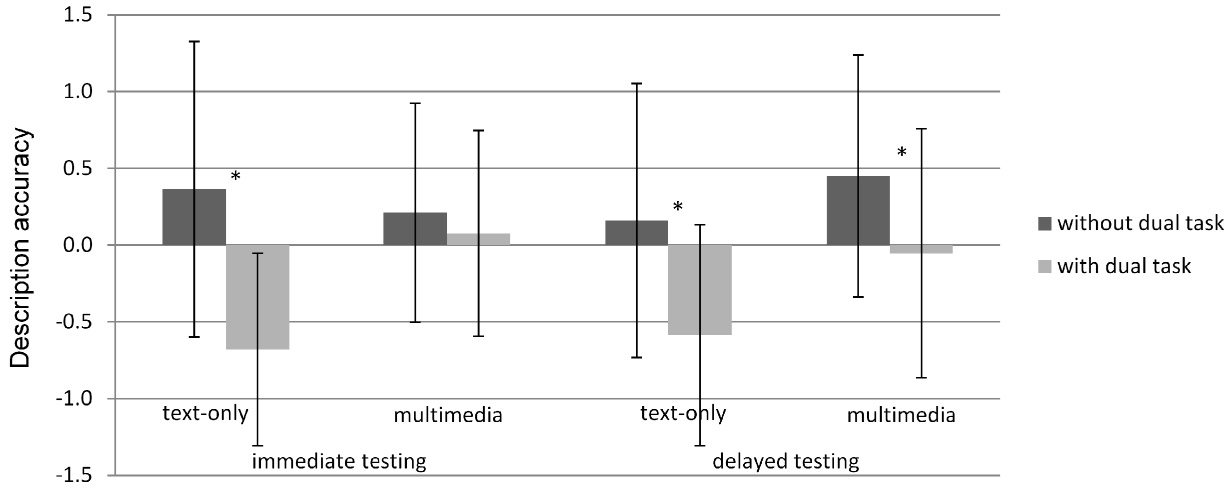

Figure 11. Means and standard deviations for the text-only and multimedia condition for description accuracy as a function of dual task and time of testing.

When looking at this 3-way interaction from a different point of view (see Figure 12), the Bonferroni adjusted pairwise comparisons showed that for both immediate and delayed testing there was a multimedia effect only for participants who learned with dual task (immediate: without dual task: p = .51, η_p^2 < .01; text-only: M = .37, SE = .17; multimedia: M = .21, SE = .17; with dual task: p < .01, η_p^2 = .11; text-only: M = -.68, SE = .17; multimedia: M = .08, SE = .17; delayed: without dual task: p = .25, η_p^2 = .02; text-only: M = .16, SE = .18; multimedia: M = .45, SE = .18; with dual task: p = .04, η_p^2 = .05; text-only: M = -.59, SE = .18; multimedia: M = -.05, SE = .18).

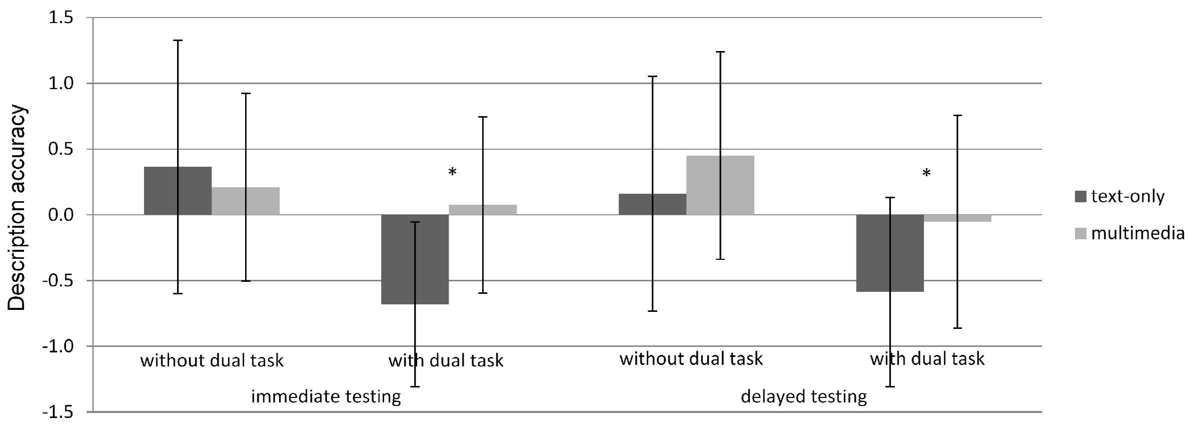

Figure 12. Means and standard deviations for the with dual task and without dual task conditions for description accuracy as a function of presentation format and time of testing.

Table 5

Means and Standard Deviations for Performance Accuracy, Description Accuracy as a Function of Dual Task, Presentation Format, and Time of Testing

	Dual task	Without dual task		With dual task	
	Presentation format	Text-only	Multimedia	Text-only	Multimedia
	Time of testing	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Performance accuracy (proportion correct)	Immediately after learning	.05 (.70)	.60 (.56)	-1.01 (.51)	.33 (.67)
	One week later	.18 (.83)	.38 (.65)	-.80 (.67)	.21 (.65)
Description accuracy (proportion correct)	Immediately after learning	.37 (.96)	.21 (.71)	-.68 (.63)	.08 (.67)
	One week later	.16 (.89)	.45 (.79)	-.59 (.72)	-.05 (.81)

4.3 Discussion

The aim of the present study was to test whether pictures are computational offloading by investigating the involvement of the visuo-spatial sketchpad when learning with text only or with multimedia in procedural tasks that are assumed to trigger imagery, taking limitations from the first two experiments into account. The procedural learning tasks conveyed actions required to perform first aid and were presented with text only or with text and pictures. The dual task approach (i.e., learning with vs. without dual task) was used to assess the involvement of the visuo-spatial sketchpad during learning. Performance of the first aid tasks was measured by performance accuracy, focusing on procedural-motor/implicit knowledge, and by description accuracy, focusing on declarative/explicit knowledge. Learning outcomes were measured immediately after learning and after one week. CTML would predict that the visuo-spatial sketchpad is more involved when learning with multimedia than when learning with text only, as the visuo-spatial sketchpad is assumed to be involved in picture processing and CTML does not take contemporary research concerning visuo-spatial text processing into account. In contrast, the account of computational offloading would predict that the visuo-spatial sketchpad is more involved when learning with text only than when learning with multimedia, as imagery requires more mental resources than visuo-spatial perception.

The results showed a multimedia effect for both learning outcome measures. However, for performance accuracy, the results also showed that performance decreased when learning with text only due to performing the dual task, whereas performance was unaffected when learning with multimedia. Accordingly, the dual task interfered to a large extent with information processing during learning with text only, but did not interfere with information processing during learning with multimedia. This implies that the visuo-spatial sketchpad was highly involved when learning with text only, but not when learning with multimedia. These results are in contrast to previous research by Gyselinck et al. (2002), Kruley et al. (1994), and Brunyé et al. (2006).

A possible explanation for this discrepancy concerning learning with text only is that their tasks and my tasks differ in the degree to which the tasks trigger imagery. The tasks of Gyselinck et al. (2002), Kruley et al. (1994), and Brunyé et al. (2006) might not have triggered imagery, as their tasks did not depict actions or did not depict actor's body parts required to perform the actions, which is assumed to trigger imagery (e.g., Woolfolk et al., 1985; Weinberg et al., 1980; Epstein, 1980). In contrast, my tasks appear to have triggered mental imagery as they depicted the actor who performed actions on objects (i.e., victim). Taken together, this suggests that the role of the visuo-spatial sketchpad is different for learning with text only when tasks trigger imagery compared to tasks that do not trigger imagery. Also, a possible explanation for this discrepancy concerning learning with multimedia, could be that the presentation of pictures in Experiment 2 substituted the need for mental imagery and therefore strongly reduced the amount of cognitive resources required to understand the task.

Furthermore, the results for performance accuracy showed that the multimedia effect decreased over time. The same pattern was found for description accuracy, but only for participants who learned with dual task — no multimedia effect was found for participants learning without dual task. The effects of time of testing in this study have to be interpreted with care. Firstly, participants were tested twice. According to the testing effect, retrieving information from memory positively influences memory for the task at a later time point (Carrier & Pashler, 1992) and therefore might also have affected performance after one week. Secondly, participants who learned with text only saw a picture from the learning material during the description post-test. As participants who learned with multimedia already saw the picture during the learning phase, this picture may have positively influenced memory for the task only for participants who learned with text only, and therefore may have affected their performance after one week. If this explanation is correct, this also shows that pictures are beneficial to learning. In this study, it is impossible to assess how these issues affected the pattern of results in each experimental condition.

Finally, the results for performance accuracy and description accuracy do not show exactly the same pattern of results. For example, the results for description accuracy did not show a multimedia effect for participants who learned without dual task. A possible explanation for this missing multimedia effect is that, when studying procedural tasks that convey actions, pictures are suitable to convey procedural-motor/implicit knowledge, as gauged by performance accuracy, however are less suitable to convey declarative/explicit knowledge, as gauged by description accuracy.

Taken together, the results seem to support the account of computational offloading best, as pictures in procedural tasks that trigger imagery by describing the position and movement of an actor’s body parts required to perform actions omit the need to engage in imagery and therewith reduce the cognitive effort that is required to understand the learning material.

5 General Discussion

In the field of multimedia learning, the multimedia effect is a well-established effect. It states that adding pictures to text is beneficial to learning. Although many empirical studies have shown *that* pictures are beneficial to learning, it is still unclear *why* they are beneficial to learning. In this dissertation, two theoretical accounts that explain the multimedia effect have been contrasted to see which account explains the multimedia effect best: the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2005b) and the account of computational offloading (Larkin & Simon, 1987). CTML posits that when learning with text and pictures, relevant information is selected from both sources and that this information is processed in working memory, then integrated with prior knowledge, and finally stored in an integrated mental representation in long-term memory. The multimedia effect is explained by arguing that learning from text and pictures yields a richer mental representation than learning from text only. In contrast, the account of computational offloading posits that imagery is required when learning with text only, whereas visuo-spatial perception can be used instead of imagery when learning from text and pictures. More cognitive resources are assumed to be required for imagery than for visuo-spatial perception. The multimedia effect is explained by arguing that information processing in working memory is facilitated when learning with text and pictures compared to learning with text only.

CTML and the account of computational offloading differ in three important ways. These differences were used to derive the hypotheses in this dissertation and concern the role of visuo-spatial information and the role of working memory during learning. In particular, they differ in 1) whether the beneficial effect of pictures depends on the amount of visuo-spatial information processing that is required by a certain type of task, 2) whether information processing in working memory differs between types of tasks, and 3) whether information processing in working memory differs between processing text only and processing text and pictures. The concept of task type, which concerns the use of visuo-spatial information in different types of tasks (i.e., conceptual, causal, and procedural) was used to investigate the role of visuo-spatial information on the multimedia effect; the concept of working memory, which concerns information processing in several working memory subsystems (i.e., the central executive, phonological loop, and visuo-spatial sketchpad) was used to investigate the role of information processing on the multimedia effect. These concepts were discussed to be able to explain the hypotheses derived from these theoretical accounts and to be able to investigate the differences between CTML and the account of computational offloading. The hypotheses that were derived were all tested in this dissertation. A summary of the results is provided next.

5.1 Summary of the Results and Interpretation

In Experiment 1, the role of visuo-spatial information and thus the effect of task type on the multimedia effect was investigated. Both theoretical accounts predicted that there would be a multimedia effect for conceptual, causal, and procedural tasks. However, CTML predicted that the size of the multimedia effect would be the same for all task types, whereas the account of computational offloading predicted that the multimedia effect would be larger in causal and procedural tasks than in conceptual tasks. The results seemed to support the account of computational offloading best, as the multimedia effect differed between task types. However, unexpectedly, the multimedia effect was larger in procedural tasks than in causal tasks for recall learning outcome measures and larger in conceptual tasks than in procedural tasks for transfer learning outcome measures.

In Experiment 2, the role of visuo-spatial information on working memory involvement during learning was investigated. CTML predicted that the central executive, the phonological loop, and the visuo-spatial sketchpad would be involved when learning with multimedia and that there would be no differences in involvement between conceptual, causal, and procedural tasks. In contrast, the account of computational offloading predicted that the involvement of the visuo-spatial sketchpad would be higher in causal and procedural tasks than in conceptual tasks. No differences between task types were predicted concerning the involvement of the central executive and phonological loop. Although the interference between the dual task and the learning task was the same between task types — which would support CTML — there was, unexpectedly, in most cases no interference between the dual task and the learning tasks at all. Only performing a verbal dual task that loaded the central executive interfered with the learning tasks when considering recall learning outcome measures.

In Experiment 3, the account of computational offloading was directly tested by comparing the involvement of the visuo-spatial part of working memory (i.e., visuo-spatial sketchpad) when learning with text only or with multimedia in procedural tasks that were assumed to trigger imagery. CTML predicted that the visuo-spatial sketchpad would be more involved when learning with multimedia than when learning with text only, whereas the account of computational offloading predicted the opposite. The results seemed to support the account of computational offloading as providing pictures reduced the cognitive effort required to understand the learning material, which was indicated by the higher interference in the visuo-spatial sketchpad when learning with text only than when learning with text and pictures.

Taken together, the results did not unanimously support CTML or the account of computational offloading. Some results supported the account of computational offloading best, such as the results that the multimedia effect differed between task types in Experiment 1 and pictures seemed to reduce the cognitive effort during learning in Experiment 3. This implies that pictures could be beneficial to learning because they facilitate information processing during learning. However, other results supported CTML best, such as the results that the multimedia effect in Experiment 1 was equally large for conceptual and causal tasks and that working memory involvement in Experiment 2 did not differ between task types. However, it should be noted that the results supporting the CTML were null effects. Unfortunately, the results did not unequivocally support any of the theoretical accounts. Important factors that might have influenced the pattern of results are discussed next.

5.2 Factors Influencing the Pattern of Results

5.2.1 Factors affecting imagery during learning

According to the account of computational offloading, imagery plays an important role during learning with text only, as mental images in working memory facilitate constructing a mental representation of the learning content in long-term memory. In the introduction, it was described that imagery involves generating or recalling mental images from long-term memory, which reflect visuo-spatial information from the real world (Glasgow & Papadias, 1995). Based on this assumption, it was argued that imagery is especially triggered when visuo-spatial information is important for understanding the learning material. Therefore, from the computational offloading point of view, the multimedia effect was expected to be larger in causal and procedural tasks than in conceptual tasks, as visuo-spatial information reflected visuo-spatial information in the real world in causal and procedural pictures, whereas space was used in a metaphorical way in conceptual pictures. However, based on the results of this dissertation, imagery that is triggered by the presentation of visuo-spatial information reflecting visuo-spatial information in the real world does not seem to be sufficient to explain differences between task types: although causal and procedural tasks both convey visuo-spatial information in a similar way, the multimedia effect in Experiment 1 was larger in procedural tasks than in causal tasks for recall learning outcome measures. Also, previous research showed differences between causal and procedural tasks by demonstrating that instructing participants to apply imagery enhanced learning for causal tasks but not for procedural tasks (e.g., Epstein, 1980; Hegarty et al., 2003a; Hegarty et al., 2003b; Weinberg et al., 1980; Woolfolk et al., 1985). Thus, is could be argued that procedural tasks automatically trigger imagery whereas causal tasks do not, which cannot be explained by the amount of visuo-spatial information that reflects visuo-spatial information in the real world, as causal and procedural tasks were suggested to be similar in that respect.

Another difference between causal and procedural tasks is that in some procedural tasks action-related words (i.e., nouns, verbs, adjectives) are used. Interestingly, these words seem to automatically trigger mental imagery (Fischer & Zwaan, 2008). Also, in some procedural tasks, the interaction between a body part and an object (e.g., a picture of a hand grabbing a glass) is described/depicted, which also seems to automatically trigger the activation of mirror neurons in the motor system (Fogassi & Ferrari, 2010; Rizzolatti & Craighero, 2004; see also Van Gog, Paas, Marcus, Ayres, & Sweller, 2009), which has been assumed to be involved in motor imagery (Kosslyn, Ganis, & Thompson, 2001). These differences might explain why instructing students to apply imagery has been shown to improve learning in causal tasks but not in procedural tasks (Weinberg et al., 1980; Woolfolk et al., 1985). Also, these differences could explain the high involvement of the visuo-spatial sketchpad in Experiment 3 when learning with text only compared to learning with text and pictures, as procedural tasks in this experiment conveyed actions. However, these two differences do not explain why the multimedia effect was larger in procedural tasks than causal tasks in Experiment 1, as in contrast to the procedural tasks in Experiment 3 these procedural tasks conveyed states and did not contain actions related words, nor did they depict interactions between body parts and objects.

An alternative explanation, as discussed after Experiment 1, might be that procedural tasks might have been more complex than causal tasks, because in Experiment 1, procedural tasks required imagery in three dimensions (e.g., the rope went behind a piece of wood), whereas causal tasks required imagery in two dimensions (e.g., the stone rolled down the hill).

If imagery is indeed more effortful with 3D than 2D space, in Experiment 1, imagery in procedural tasks can be assumed to be more effortful than imagery in causal tasks. The idea that imagery in 3D space is effortful is supported by the finding from Experiment 3 that studying procedural first aid tasks that required 3D information processing (e.g., the actor stood behind the victim) imposes a high load on the visuo-spatial sketchpad when learning with text only. Whether the results of Experiment 2 also support this explanation cannot be determined, as in Experiment 2, no text-only condition was included.

If this explanation is correct, it could be posited that both CTML and the account of computational offloading can explain the multimedia effect by arguing that pictures in most cases elicit a multimedia effect due to dual coding, but that in tasks that require imagery in 3D space, pictures in addition enable computational offloading and therefore elicit an even larger multimedia effect (see also Section 5.5.1). Future research, however, is needed to explicitly test the role of imagery and the load that is imposed on working memory when studying causal and procedural multimedia learning tasks that involve 2D versus 3D use of space.

5.2.2 Factors affecting the assessment of working memory involvement during learning

In this dissertation, the role of working memory has been assessed using previously applied dual tasks (i.e., random articulatory suppression task, random foot tapping task, articulatory suppression task, and foot tapping task). However, the dual tasks in Experiment 2 did not influence performance in the learning task, especially for the random foot tapping task, articulatory suppression task, and foot tapping task (see Schüler et al., 2011 for a review).

A possible explanation of why the load on the central executive when the dual task involved a visuo-spatial component (i.e., random foot tapping task) was low, is that participants' random tapping behaviour was restricted by the number of pedals, and participants could therefore not tap a wrong pedal. In contrast, when performing the central executive dual task with a verbal component (i.e., random articulatory suppression task), participants were required to inhibit responses of numbers other than one, two, three, and four, implying that they could name a wrong number such as five. Inhibiting responses requires additional information processing in the central executive (Baddeley, 1986). Therefore, the random articulatory suppression task might have required additional central executive resources compared to the random foot tapping task. This means that the load on the central executive might have been lower when performing the random foot tapping task than when performing the random articulatory suppression task. This load — when considering the results — may not have been high enough to cause interference with the learning task.

A possible explanation for why the articulatory suppression task did not affect information processing in the phonological loop could be that the phonological loop is, according to Baddeley (2006), not highly involved when skilled readers read a text, as the phonological loop does not seem to be influenced when language processing is a habit. As participants were university students, it can be assumed that they were highly skilled readers. Therefore, reading can be considered as a routine task for the participants, implying that processing verbal information from the text caused a low load in the phonological loop and therefore did not interfere with the learning task. However, it should be noted that other empirical results concerning the role of the phonological loop during multimedia learning (Brunyé et al., 2006; Gyselinck et al., 2002; Kruley et al., 1994; see Section 1.6.3.2) do not support Baddeley's argument, as there was interference between the dual task that loaded the phonological loop and the learning tasks. A difference between the articulatory suppression task used in previous research and the articulatory suppression task used in Experiment 2 was that in the present study participants uttered numbers (i.e., one, two, three, four), whereas participants in previous studies uttered syllables (i.e., ba, be, bi, bo [Brunyé et al., 2006] or ba, be, bi, bo, bu [Gyselinck et al., 2002]). Although numbers and syllables are both verbal sequences and therefore require information processing in the phonological loop (Baddeley, 1986), a possible explanation of the missing interference could be that participants were more familiar with numbers than with these syllables and thus required less information processing in the phonological loop.

A possible explanation for why the foot tapping task did not affect information processing in the visuo-spatial sketchpad is that in the learning tasks, visuo-spatial information was relevant for understanding the task, but could be extracted so easily from the pictures that processing these pictures did not overload the visuo-spatial sketchpad. This postulation is supported by the results of Experiment 3, which also showed no interference between learning with text and pictures and performing the foot tapping tasks in the visuo-spatial sketchpad.

In general, an important difference between Experiment 2 and previous research on working memory involvement during multimedia learning is that Experiment 2 only involved learning with text and pictures and did not involve text-only control conditions. This limitation was addressed in Experiment 3 by including text-only conditions in the experimental design. Future research, therefore, should focus again on the role of working memory in eliciting a multimedia effect in different task types by assessing working memory involvement in both text-only and multimedia conditions.

5.2.3 Factors affecting the expressiveness of the learning outcome measures

In Experiment 1 and 2, several dependent variables have been used to measure different types of knowledge. In accordance with theoretical assumptions (cf. Mayer, 2009b), it was relevant to distinguish between recall and transfer of information. In Experiment 1, the multimedia effect was largest in procedural tasks and equally large in conceptual and causal tasks for recall measures (i.e., free recall and recall verification accuracy), but larger in conceptual tasks than procedural tasks for transfer measures (i.e., transfer verification accuracy, and integration verification accuracy). Furthermore, in Experiment 2, the results showed interference between learning and performing a dual task that loaded the central executive with a verbal component (i.e., random articulatory suppression task) for recall measures (i.e., free recall and recall verification accuracy), but not for transfer measures (i.e., free transfer, transfer verification accuracy, and integration verification accuracy). In Experiment 3, only recall measures were used.

Surprisingly, there were several inconsistencies between dependent variables in all three experiments. In Experiment 1, the multimedia effect was equally large between conceptual and causal tasks for both recall measures; however, only for free recall, the multimedia effect for these task types was significant. Also, in Experiment 2, there were no significant results for free transfer; however, for transfer verification accuracy, performance in the visuo-spatial sketchpad condition was higher than in the verbal central executive condition and for integration verification accuracy, performance in the visuo-spatial central executive condition was higher than in the verbal central executive condition. Furthermore, in Experiment 3, there was for example a multimedia effect for performance accuracy independent of whether participants performed a dual task, whereas for description accuracy there was only a multimedia effect for participants who performed a dual task.

Based on these observations, it can be argued that certain factors affected the expressiveness of the learning outcome measures and therewith influenced whether certain effects were found. As discussed in Experiment 2, one factor affecting expressiveness is the ease of the post-test questions. This involves whether answers were easy to guess, but also whether the difficulty of questions caused a ceiling or flooring effect (i.e., when all participants, independent of condition, were able or not able to answer the question correctly). As the learning material and post-test questions in Experiment 1 and 2 were especially designed for this dissertation, it was not clear in advance whether questions were easy to guess or would cause ceiling or flooring effects. Therefore, I suggest that the item difficulty of the post-test items is assessed to decide whether these post-test questions can be used again in future research.

Another factor affecting the expressiveness of learning outcome measures, as discussed in Experiment 3, is whether the learning outcome measure is suitable to assess the type of knowledge that is constructed. For example, in Experiment 1 and 2, procedural knowledge was assessed by similar recall and transfer measures as causal and conceptual tasks; accuracy of performing the procedure was not taken into account. In contrast, in Experiment 3, knowledge concerning procedures was assessed by requiring participants to perform the procedure, and not only by requiring participants to describe the procedure — which is also sometimes used to assess learning outcome measures in procedural tasks (e.g., Arguel & Jamet, 2009; Brunyé et al., 2006). Assessing performance accuracy directly showed whether participants were able to correctly recall the procedure, whereas assessing description accuracy merely showed whether participants were able to correctly recall the description of the procedure. Therefore, in future research, it should be considered which type of learning outcome is most suitable to assess the type of knowledge that is constructed.

A third factor affecting the expressiveness of the learning outcome measures is the sensitivity of the experimental measures. In the experiments, the free recall, performance accuracy, and description accuracy measures assessed learning outcome on a very detailed level (i.e., scores could range between 0 and 30 at least and 429 at most in Experiment 1 and 2, and between 0 and 16 at least and 48 at most before standardising in Experiment 3), whereas scores using verification items were less detailed (i.e., scores ranged between 0 and 3 at least and 18 at most). Improving the sensitivity of the dependent measure reduces the level of error — which can distort experimental effects — thereby enhancing statistical power

(Lipsey, 1990). If the sensitivity was indeed lower in certain learning outcomes measures, this could for example explain why the multimedia effect for conceptual and causal tasks was not significant for recall verification accuracy, however was significant for free recall in Experiment 1. I suggest that in future research, the expressiveness of learning outcome measures is increased not only by using more suitable items and/or learning outcome measures, but also by considering the sensitivity of these measures.

A fourth factor affecting the expressiveness of the learning outcome measures is the distinction between information sources. As discussed after Experiment 2 (see Section 3.3), distinguishing between information sources is important when investigating the multimedia effect. A multimedia effect is not expected in case information is only given in the text, as participants in both text-only and multimedia conditions have access to this information. In Experiment 1, this was confirmed for all dependent variables. A large multimedia effect is assumed in case information is only given in the picture, as participants in the text-only condition did not have access to this information. In Experiment 1, this was also confirmed for all dependent variables. The multimedia effect is especially interesting in case information is given in both text and pictures, as participants in both text-only and multimedia conditions have access to this information, but participants in the latter condition are still assumed to remember more information. In Experiment 1, a multimedia effect for text-picture questions was only found for recall measures. When the distinction between information sources is not made, interpreting a multimedia effect is difficult, as the effect could be caused by picture questions that participants in the text-only condition were not able to answer. In contrast, interpreting a missing multimedia effect is also difficult, as the missing effect could be caused by text questions that involve information that is not conveyed by the picture.

A final, albeit rather hypothetical, factor affecting the expressiveness of the learning outcome measures might be the ease of transforming a mental representation into an external representation during recall of information. For example, participants in Experiment 1 and 2 were required to write down and draw everything they remembered. Participants who learned with text and pictures could draw and describe parts of the picture that they had seen. In contrast, participants who had learned with text only had to construct an image based on the information stored in memory during recall in case they had not applied imagery during learning. As participants had a maximum of five minutes for writing and for drawing, constructing an external representation might therefore have required more time, enabling them to write or draw less in the same amount of time, thereby evoking lower learning outcomes. As procedural tasks concerned 3D space (see Section 1.5.3), this argument could especially apply for procedural tasks, and could therefore be an alternative explanation for why the multimedia effect in Experiment 1 was larger in procedural tasks than in causal and conceptual tasks. However, as discussed after Experiment 2 (see Section 3.3), the use of 3D space might have affected difficulty of both the learning task and the writing and drawing tasks. As the interplay between difficulty of the task and difficulty of constructing an external representation cannot be disentangled, the validity of this argument can unfortunately not be assessed based on the experiments and results in this dissertation and should therefore be addressed in future research.

5.3 Limitations

Several limitations of the experiments in this dissertation have already been discussed. One limitation of Experiment 1 and 2 is that the difficulty of both post-test questions might have influenced the pattern of results (see Section 3.3 and 5.2.3), and that the interplay between the difficulty of both post-test questions and learning tasks can unfortunately not be unravelled. Another limitation of Experiment 1 and 2 was that learning outcome measures for procedural tasks did not involve performance accuracy; however, this limitation was addressed in Experiment 3 by requiring participants not only to describe the procedure, but also to perform the procedure. A limitation of Experiment 2 was that no text-only conditions were included (see Section 5.2.2).

A further limitation of Experiment 1 and 2 was that fictitious learning material was used and that therefore the external validity of the results can be questioned. However, the advantage of using fictitious learning material was that the comparability between tasks was controllable, which increased the internal validity of the experiments. I believe that using fictitious learning materials when comparing task types enabled me to optimise the trade-off between internal validity (i.e., comparability between task types) and external validity (i.e., comparability with multimedia learning tasks in non-experimental situations). In Experiment 3, the comparability between task types was not an issue, which enabled me to use non-fictitious learning materials and thus increase the external validity.

A theoretical limitation is that Larkin and Simon (1987), who proposed the account of computational offloading, based their ideas on text only processing and picture only processing, whereas in this dissertation text only was compared with text and picture processing. As presenting text and pictures could trigger processes that are not possible when learning with only once source, such as integration of information from text and picture, the interplay between text and pictures can affect information processing during learning. For example, in this dissertation, it was assumed that texts trigger imagery whereas pictures trigger visuo-spatial perception, and that visuo-spatial perception substitutes imagery. However, when learning with text and pictures, it is also possible that both imagery and visuo-spatial perception are used to understand the learning material; it is even possible that participants only apply imagery in case they do not look at the picture. As the results do not always support the account of computational offloading, the interplay between text and pictures might have caused the diffuse results. Future research that is in line with Larkin and Simon's original research should be conducted to investigate whether their postulations also apply when combining text and pictures.

Another theoretical limitation of this dissertation was that the explanations of why pictures are beneficial to learning (i.e., dual coding vs. computational offloading) were tested indirectly: to be able to contrast the two theoretical accounts, the concepts of task type and working memory were used. To assess dual coding, differences in the size of the multimedia effect between task types were used. A disadvantage of this indirect way of testing is that many assumptions had to be made concerning the relevance of visuo-spatial information in different task types and that it is still unclear whether these assumptions are correct. A more direct test would be to assess what types of mental representations are built based on the learning material. However, it should be noted that this could be challenging as any measure is affected by the underlying phenomenon of interest (i.e., the mental representation), but also by how this phenomenon is translated into observable effects (Kosslyn, 1978). Therefore, it could be difficult to know how the outcome measure is affected by the underlying phenomenon and how it is affected by its translation. Alternatively, when continuing to consider the role of visuo-spatial information, the type of visuo-spatial information (e.g., 2D vs. 3D) and the influence on the multimedia effect could be assessed systematically (as already suggested in Section 5.2.1). Furthermore, to assess computational offloading, the interference in the visuo-spatial sketchpad between the learning task and a visuo-spatial dual task was gauged. A disadvantage of this indirect way of testing is that many assumption concerning which processes take place and how much load they impose on working memory had to be made. Another disadvantage is that it cannot be unequivocally derived from the results how much load was imposed by which mental process. A more direct test would be to record process data that provides information about whether and when imagery was applied. A possible source of process data could be think-aloud verbal protocols that concern subjects' reports on their use of imagery during the learning task (cf. Russo, Johnson, & Stephens, 1989). Another source of process data could be eye movements on blank space (i.e., blank screen paradigm; Altmann, 2004), which are performed during imagery and are assumed to reflect the spatial structure of the underlying mental representation (see also Eitel, Scheiter, Schüler, Nyström, & Holmqvist, in press).

5.4 Strengths

Apart from these limitations, the studies reported in this dissertation had several strengths. One strength of this dissertation is that in all three experiments, results always depended on more than one learning task. Moreover, in verification measures, several questions were included. With only one learning task or only one question, unique characteristics that are not of interest and are unrelated to the characteristic being measured (e.g., the number of different colours in the picture) can cause fluctuations in the measure and can therefore cause unreliability (Lipsey, 1990). By using several learning tasks and several questions for verification accuracy measures, the influence of unique characteristics was reduced and therefore the reliability of the results increased. Also, it preserved me from drawing conclusions that applied only to one specific measure and therefore from unjustly generalising to other tasks and dependent measures. This in turn increased the reliability of the conclusions that were drawn in this dissertation.

A second strength of this dissertation is that in Experiment 1 and 2, different information sources were considered. Similar to including several types of learning outcome measures and at least two measures within each type of learning outcome measure, distinguishing between information sources preserved me from unjustly generalising the multimedia effect and therefore enabled me to draw more accurate and detailed conclusions concerning the multimedia effect (see also Section 5.2.3).

A third strength of this dissertation is that in Experiment 1 and 2 the effect of task type was investigated systematically. Reigeluth and Stein (1983) already distinguished between conceptual, causal, and procedural tasks and Scaife and Rogers (1996) already argued that “the value of different graphical representations [i.e., pictures] … cannot be assessed adequately from our intuitions. To be effective a number of interdependent factors need to be considered, such as … the type of task.” (p. 186). Nevertheless, the multimedia effect in different task types has, to my knowledge, not been investigated systematically. As the results of Experiment 1 showed that it is indeed relevant to distinguish between task types, I suggest that future research will continue to distinguish between task types or at least consider the role of visuo-spatial information.

A fourth strength of this dissertation is that in Experiment 2 and 3, the dual task approach was used to investigate working memory involvement during multimedia learning. So far, only a few studies have used this approach with multimedia learning materials (i.e., Brunyé et al., 2006; Gyselinck et al., 2002; Kruley et al., 1994). Schüler et al. (2011) argued that this might be because the dual task approach is difficult to implement, as 1) multimedia learning material is more complex than stimuli in basic cognitive psychology experiments, 2) it is more difficult to generate multiple instances of the instructional materials, and 3) features of experimental materials are more difficult to control systematically. To be able to control learning material as systematically as possible, I created fictitious multimedia learning materials. I tried to make the tasks as similar as possible concerning the length of the texts, the length of sentences, the number of propositions in text/picture/both, whether texts were easy to follow, whether texts were highly concrete, whether texts were engaging, and the type of pictures (i.e., coloured line drawings). By controlling these aspects, it was possible to create multiple instances of the instructional material. As the complexity of stimuli is inherent to multimedia learning material, the learning tasks in this dissertation were similar to common learning tasks used in experimental research (e.g., Mayer, 2009b) and used in school books (e.g., Henny, 2006), to be able to maintain external validity and comparability. Taken together, the controlled learning tasks and multiple instances enabled me to apply the dual task approach appropriately.

5.5 Implications

Despite the possible limitations and because of the denoted strengths, the results of the reported experiments have important theoretical and practical implications.

5.5.1 Theoretical implications

The results of the experiments reported in this dissertation did not unanimously support CTML or the account of computational offloading. In Section 5.2.1, it was argued that the multimedia effect could be explained by arguing that pictures in most cases elicit a multimedia effect due to dual coding, but in tasks that require imagery in 3D space, pictures in addition enable computational offloading and therefore elicit an even larger multimedia effect. A possible theoretical implication would be that CTML and the account of computational offloading should not been seen as two distinct and independent views on how a multimedia effect is elicited. Instead, it might be relevant to combine these views in one model that enables explaining the multimedia effect by both dual coding and computational offloading, depending on whether imagery in 2D or 3D space is triggered. Mayer's model already includes the idea of dual coding (i.e., verbal and pictorial mental models) and of visuo-spatial perception (i.e., observing pictures with the eyes). Next, I argue how the model can be adapted so that it incorporates the process of imagery and therewith predicts the idea of computational offloading.

When applying imagery, from a CTML point of view, words are observed initially, then these words are processed in the word sound base, and after that these words are processed in the visual image base. This process can be visualised by connecting Mayer's component 'words' with his component 'eyes', then connecting 'eye' with 'word sound base' — which is not part of his original model — and connecting 'word sound base' with 'visual image base' (see bold arrows in Figure 13). If the learning task triggers imagery in 2D space, the transition from the word sound base to the visual image base is assumed to be effortful. However, this transition is assumed to be even more effortful when imagery in 3D space is triggered.

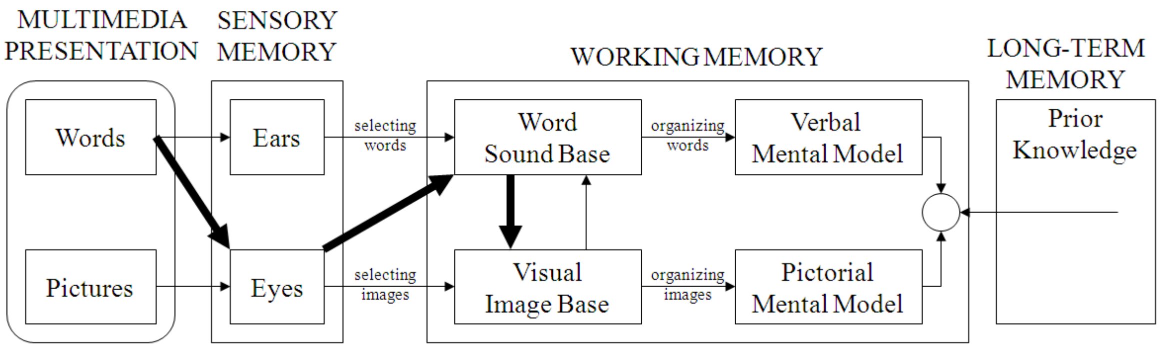

Figure 13. CTML adapted to include the process of imagery.

By visualising this process, it becomes clear how the model can explain why pictures are computational offloading: the visuo-spatial perception route from ‘pictures’ to ‘eyes’ to ‘visual image base’ seems to be more direct — implying that less information processing in working memory is required and pictures are therefore computational offloading — than the indirect imagery route from ‘words’ to ‘eyes’ to ‘word sound base’ to ‘visual image base’. When 3D imagery is involved, the effort required for the transition from ‘word sound base’ to ‘visual image base’ is even larger, implying that the difference in effort between visuo-spatial perception and imagery is even larger and pictures are even more computational offloading.

This theoretical implication seems to be relevant also from other theoretical points of view (e.g., Rummer et al., 2008; Schüler, 2010). However, it should be noted that the validity of this suggestion depends on the research that was proposed (see Section 5.2.1) concerning the role of imagery and the load that is imposed on working memory in causal and procedural multimedia learning tasks that involve 2D versus 3D use of space.

5.5.2 Practical implications

In Experiment 1, pictures seemed to be helpful only when a post-test question could be answered using information provided in the picture (i.e., text-picture questions and picture questions). In other words, if the relevant information is only given in the text, pictures do not seem to be beneficial to learning (cf. Levie & Lentz, 1982). This statement may seem trivial at first sight; however, it supports the view that the multimedia effect is a cognitive and not a motivational effect. If the multimedia effect was due to a motivational mechanism, where presenting a picture yields higher motivation to learn and engage in the task, then this higher motivation and engagement should positively influence learning as a whole. However, as the multimedia effect is limited to the information that is conveyed through pictures, the effect is instead based on a cognitive advantage because this information is more accessible (cf. CTML) or more easily processed (cf. account of computational offloading). This argument is supported by Carney and Levin (2002) who showed that decorative pictures (i.e., pictures that decorate the page but bear little or no relationship to the text content) are not beneficial to learning, whereas for example representational pictures (i.e., pictures that mirror part or all of the text content) and interpretational pictures (i.e., pictures that help to clarify difficult text) do support learning. This implies that pictures should convey information relevant to the learning task, as pictures do not foster learning by merely improving motivation.

Also, in Experiment 1 and 3, even though this still needs to be confirmed by replication of the experiment, the beneficial effect of pictures seems to depend on the type of learning outcomes (see also Section 5.2.3). Therefore, teachers and instructional designers should consider which type of learning is required for a given task. In Experiment 1, the multimedia effect seemed to be largest in procedural tasks when learning outcomes measures concerned recall of information, whereas the multimedia effect seemed to be largest in conceptual tasks when learning outcomes measures concerned transfer of information. If procedural content is conveyed that focuses on actions, performance accuracy seems to be a more reliable learning outcome measure than description accuracy.

Based on Experiment 3, pictures seem to be beneficial to learning by reducing the cognitive effort that is required to understand the learning material. By reducing the required cognitive effort for processing one part of the learning task, more effort can be spent on other parts of the learning task. Even though Carney and Levin (2002) argued that presenting pictures is superfluous when text elicits mental images in students, the results from this study suggest that teachers and instructional designers should consider using pictures also when a task triggers imagery to facilitate understanding of the learning material.

5.6 Concluding Remarks

The research presented in this dissertation was, to my knowledge, the first attempt to contrast CTML (i.e., the view that pictures are beneficial to learning because they allow enhanced information storage) and the account of computational offloading (i.e., the view that pictures are beneficial to learning because pictures facilitate information processing during learning) by using controlled material (i.e., different task types) and methods from basic cognitive research (i.e., dual task approach). As the results of this dissertation do not seem to unanimously support CTML or the account of computational offloading, it was suggested that a view that combines both approaches might be theoretically relevant. Nevertheless, I suggest that future research continues to address the question why pictures are beneficial to learning, taking limitations and strengths from the series of experiments reported in this dissertation into account.

6 Summary

In the field of multimedia learning, the multimedia effect, which states that adding pictures to text is beneficial to learning, is a well-established effect. Although many empirical studies have shown *that* pictures are beneficial to learning, it is still unclear *why* they are beneficial to learning. In this dissertation, two theoretical accounts that explain the multimedia effect have been contrasted to see which account explains the multimedia effect best: the Cognitive Theory of Multimedia Learning (CTML) and the account of computational offloading. CTML explains the multimedia effect by arguing that learning from text and pictures yields a richer mental representation than learning from text only. In contrast, the account of computational offloading explains the multimedia effect by arguing that information processing in working memory is facilitated when learning with text and pictures compared to learning with text only.

In this dissertation, three experiments are reported that tested 1) whether the beneficial effect of pictures depends on the amount of visuo-spatial information processing that is required by a certain type of task (i.e., conceptual, causal, procedural), 2) whether information processing in working memory differs between types of tasks, and 3) whether information processing in working memory differs between processing text only or processing text and pictures. The results did not unanimously support any of the theoretical accounts. Therefore, it was argued that a theoretical account that combines both explanations of the multimedia effect, stating that pictures yield a richer mental representation and that pictures are under certain circumstances computational offloading, seems to be most likely. However, it is noted that further research is needed that continues to address the question why pictures are beneficial to learning, taking the reported limitations and strengths into account.

7 Zusammenfassung

Im Forschungsbereich des multimedialen Lernens wurde der Multimediaeffekt, d.h. der Befund, dass die Darbietung von Texten zusammen mit Bildern lernförderlich ist, empirisch vielfach bestätigt. Trotz dieser eindeutigen Befundlage ist jedoch unklar, worauf der Multimediaeffekt zurückzuführen ist. In der vorliegenden Dissertation wurden daher zwei theoretische Erklärungen für den Multimediaeffekt kontrastiert, nämlich die Cognitive Theory of Multimedia Learning (CTML) sowie die Annahme des Computational Offloading. Nach der CTML geht der Multimediaeffekt darauf zurück, dass Text-Bilddarbietungen zu reichhaltigeren mentalen Repräsentationen führen als die alleinige Textdarbietung. Im Gegensatz dazu erklärt die Annahme des Computational Offloading den Multimediaeffekt durch eine erleichterte Informationsverarbeitung im Arbeitsgedächtnis, wenn Texte zusammen mit Bildern dargeboten werden.

Im Rahmen der vorliegenden Dissertation wurden drei Experimente durchgeführt, die testeten (1) ob der Multimediaeffekt vom Ausmaß an visuell-räumlicher Information, welches verschiedenen Aufgabentypen (d.h., konzeptuellen, kausalen und prozeduralen Aufgaben) inhärent ist, abhängt (2) ob sich die Informationsverarbeitung im Arbeitsgedächtnis je nach Aufgabentyp verändert und (3) ob sich die Informationsverarbeitung von Texten und Bildern im Vergleich zur reinen Textverarbeitung im Arbeitsgedächtnis unterscheidet. Die Ergebnisse der Experimente konnten keine der beiden theoretischen Annahmen eindeutig bestätigen. Daher wird angenommen, dass eine Kombination aus beiden Erklärungen den Multimediaeffekt am besten erklärt, nämlich dass Text-Bilddarbietungen zu reichhaltigeren mentalen Repräsentationen führen und dass unter bestimmten Bedingungen Bilder auch die Informationsverarbeitung erleichtern können. Weitere Forschung zur Frage, warum Bilder das Lernen unterstützen, ist jedoch notwendig. Zukünftige Arbeiten sollten dabei die Stärken und Schwächen der vorliegenden Dissertation berücksichtigen.

8 References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education, 33,* 131-152.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: the ‘blank screen paradigm’. *Cognition*, *93,* B79-B87.

Andrade, J. (2001). *Working memory in perspective.* Hove: Psychology Press.

Anglin, G. J., Vaez, H., & Cunningham, K. L. (2004). Visual representations and learning: The role of static and animated graphics. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 865-916). Mahwah, NJ: Erlbaum Associates.

Arguel, A., & Jamet, E. (2009). Using video and static pictures to improve learning of procedural contents. *Computers in Human Behavior*, *25,* 354-359.

Baddeley, A. D. (1986). *Working memory.* Oxford, UK: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science, 255,* 556-559.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology, 49A(1),* 5-28.

Baddeley, A. D. (1999). *Essentials of human memory.* Hove: Psychology Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Baddeley, A. D. (2006). Working memory: An overview. In S. J. Pickering (Ed.), *Working Memory and Education* (pp. 1-31). San Diego: Elsevier.

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford: University Press.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1-29.

Baddeley, A. D., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33,* 439-454.

Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology, 51A*, 819-852.

Bovair, S., & Kieras, D. E. (1984). A guide to propositional analysis for research on technical prose. In B. K. Britton & J. B. Black (Eds.), *Understanding expository text* (pp. 315-362). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brunyé, T. T., Taylor, H. A., Rapp, D. N., & Spiro, A. B. (2006). Learning procedures: The role of working memory in multimedia learning experiences. *Applied Cognitive Psychology, 20,* 917-940.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review, 14,* 5-26.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633-642.

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition, 30,* 1086-1095.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor anova designs. *Educational and Psychological Measurement, 33,* 107-112.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

della Sala, S., Gray, C., Baddeley, A. D., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia, 37,* 1189-1199.

Denis, M. (1989). *Image & Cognition.* NY: Harvester Wheatsheaf.

de Vega, M., Cocude, M., Denis, M., Rodrigo, M. J., & Zimmer, H. D. (2001). The interface between language and visuo-spacial representation. In M. Denis, B. H. Logie, C. Cornoldi, M. de Vega & J. Engelkamp (Eds.), *Imagery, language and visuo-spatial thinking* (pp. 109-136). Hove: Psychology Press.

Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (in press). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction*.

Epstein, M. L. (1980). The relationship of mental imagery and mental rehearsal to performance of a motor task. *Journal of Sport Psychology, 2,* 211-220.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology, 61,* 825-850.

Fogassi, L., & Ferrari, P. F. (2011). Mirror systems. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*, 22-38.

Gathercole, S., & Baddeley, A. D. (1993). *Working memory and language processing.* Hove: Lawrence Erlbaum Associates.

Glasgow, J. I., & Papadias, D. (1995). Computational imagery. In J. Glasgow, N. H. Narayanan, & B. Chandrasekaran (Eds.), *Diagrammatic reasoning: Cognitive and computational perspectives* (pp. 435-480), Cambridge, MA: MIT-Press.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models*. Journal of Memory and Language, 31,* 129-151.

Glenberg, A. M., Kruley, P., & Langston, W. E. (1994). Analogical process in comprehension: Simulation of a mental model. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 609-640). San Diego, CA: Academic Press.

Gyselinck, V., De Beni, R., Pazzaglia, F., Meneghetti, C., & Mondoloni, A. (2007). Working memory components and imagery instructions in the elaboration of a spatial mental model. *Psychological Research, 71,* 373-382.

Gyselinck, V., & Tardieu, H. (1999). The role of illustrations in text comprehension: What, when, for whom, and why? In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 195-218). London: Lawrence Erlbaum Associates.

Gyselinck, V., Cornoldi, C., Dubois, V., de Beni, R., & Ehrlich, M.-F. (2002). Visuospatial memory and phonological loop in learning from multimedia. *Applied Cognitive Psychology, 16*, 665-685.

Gyselinck, V., Ehrlich, M.-F., Cornoldi, C., de Beni, R., & Dubois, V. (2000). Visuospatial working memory in learning from multimedia systems. *Journal of Computer Assisted Learning*, *16*, 166-176.

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, *24*, 95-123.

Hegarty, M., & Just, M.A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*, 717-742.

Hegarty, M., Kriz, S., & Cate, C. (2003a). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction*, *21,* 325-360.

Hegarty, M., Narayanan, N. H., & Freitas, P. (2003b). Understanding machine from multimedia and hypermedia presentations. In J. Otero, A. C. Graesser, & J. Leon (Eds.), *The psychology of science text comprehension* (pp. 357-384). Mahwah, NJ: Lawrence Erlbaum Associates.

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science, 30,* 581-592.

Henny, W. (2006). *Orange Cross Manual, official manual for providing first aid* (25th ed.). Utrecht: Thieme Meulenhoff.

Hiebert, J., & Carpenter, T.P. (1992). Learning and teaching with understanding. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-97). New York: MacMillan.

Imhof, B., & Scheiter, K., & Gerjets, P. (2011). Learning about locomotion patterns from visualizations: Effects of presentation format and realism. *Computers & Education, 57,* 1961-1970.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85(5),* 363-394.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18,* 513-549.

Kosslyn, S. M. (1978). Imagery and internal representation. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorization* (pp. 217-257). Hillsdale, NJ: Erlbaum Associates.

Kosslyn, S. M. (1999). *Image and brain: The resolution of the imagery debate* (4th ed.). Cambridge, MA: MIT Press.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience, 2,* 635-642.

Kruley, P., Sciama, S. C., & Glenberg, A. M. (1994). On-line processing of textual illustrations in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition*, *22,* 261-272.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33,* 159-174.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Lawrence, B. M., Myerson, J., Oonk, H. M., & Abrams, R. A. (2001). The effects of eye and limb movements on working memory. *Memory, 9*, 433-444.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A Review of research. *Educational Technology Research and Development*, *30*, 195-232.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research* (Vol. 19). SAGE Publications, Incorporated .

Logie, R. H. (1995). *Visuo-spatial working memory.* Aberdeen: Lawrence Erlbaum Associates.

Logie, R. H., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial or central executive? In R. H. Logie & M. Denis (Eds.), *Mental images in human cognition* (pp. 105–115). Amsterdam: North-Holland Press.

Mayer, R. E. (2005a). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1-16). New York, NY: Cambridge University Press.

Mayer, R. E. (2005b). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York, NY: Cambridge University Press.

Mayer, R.E. (2009a). The promise of multimedia learning. In R. E. Mayer (Ed.), *Multimedia learning second edition* (pp. 3-27). New York, NY: Cambridge University Press.

Mayer, R.E. (2009b). Multimedia principle. In R. E. Mayer (Ed.), *Multimedia learning second edition* (pp. 223-241). New York, NY: Cambridge University Press.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*(4), 715-726.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology,* 86, 389-401.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89-115.

Michas, I. C., & Berry, D. C. (2000). Learning a procedural task: Effectiveness of multimedia presentations. *Applied Cognitive Psychology*, *14*, 555-575.

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115,* 123-142.

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development, 53,* 35-45.

Münzer, S., Seufert, T., & Brünken, R. (2009). Learning from multimedia presentations: Facilitation function of animations and spatial abilities. *Learning and Individual Differences, 19*, 481-485.

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology, 21,* 263-276.

Nam, C. S., & Pujari, A. (2005). *The role of working memory in multimedia learning.* In Proceedings of the 11th International Conference on Human-Computer Interaction (HCII '05). Las Vegas: Mira Digital Publishing. CD-ROM.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*, 255-287.

Pearson, D., de Beni, R., & Cornoldi, C. (2001). The generation, maintenance, and transformation of visuo-spatial mental images. In M. Denis, R. H. Logie, C. Cornoldi, M. de Vega, & J. Engelkamp (Eds.), *Imagery, language, and visuo-spatial thinking* (pp. 1-28). Hove: Psychology Press.

Reigeluth, C. M., & Stein, F. S. (1983). The elaboration theory of instruction. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status.* Hillsdale, NJ: Erlbaum Associates.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,* 169-192.

Rummer, R., Schweppe, J., Scheiter, K., & Gerjets, P. (2008). Lernen mit multimedia: Die kognitiven Grundlagen des Modalitätseffekt [Multimedia learning and the cognitive basis of the modality effect]. *Psychologische Rundschau, 59(2)*, 98-107.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, *17*, 759-769.

Ryle, G. (2000). *The concept of mind.* Chicago: University of Chicago Press.

Sadoski, M. (2001). Resolving the effects of concreteness on interest, comprehension, and learning important ideas from text. *Educational Psychology Review, 13,* 263-281.

Sadoski, M., & Paivio, A. (2004). A dual coding theoretical model of reading. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed.) (pp. 1329-1362). Newark, DE: International Reading Association.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal Human-Computer Studies*, *45*, 185-213.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Schüler, A. (2010). *Interferieren (geschriebene) räumliche Textinhalte mit der Bildverarbeitung im visuell-räumlichen Arbeitsgedächtnis? Testung einer erweiterten kognitiven Theorie multimedialen Lernens [Do (written) spatial text contents interfere with picture processing in visuo-spatial working memory? Testing an extended Cognitive Theory of Multimedia Learning]*. Doctoral dissertation. Tübingen: Knowledge Media Research Center.

Schüler, A., Scheiter, K., & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Educational Psychology Review, 23,* 389-411.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283,* 1657-1661.

Smyth, M. M., & Pendleton, L. R. (1990). Space and movement in working memory. *Quarterly Journal of Experimental Psychology, 42,* 291-304.

Stone, D. E., & Glock, M. E. (1981). How do young adults read directions with and without pictures? *Journal of Educational Psychology, 73*, 419-426.

Tversky, B., Heiser, J., Lozano, S., MacKenzie, R., & Morrison, J. (2008). Enriching animations. In R. K. Lowe & W. Schnotz (Eds.), *Learning with animation: Research and design implications.* New York: Cambridge University Press.

Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror-neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educational Psychology Review*, *21*, 21-30.

Waddill, P. J., & McDaniel, M. A. (1992). Pictorial enhancement of text memory: Limitations imposed by picture type and comprehension skill. *Memory & Cognition, 20(5),* 472-482.

Weinberg, R. S., Gould, D., Jackson, A., & Barnes, P. (1980). *Perceptual and Motor Skills, 50,* 663-666.

Wilcox, W. C., Merrill, M. D., & Black, H. B. (1981). Effect of teaching a conceptual hierarchy on concept classification performance. *Journal of Instructional Development, 5,* 8-13.

Woolfolk, R. L., Murphy, S. M., Gottesfeld, D., & Aitken, D. (1985). Effects of mental rehearsal of task motor activity and mental depiction of task outcome on motor skill performance. *Journal of Sport Psychology, 7,* 191-197.

Zhang, J. J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87-122.

9 List of Figures

Figure 1. Cognitive Theory of Multimedia Learning (obtained from Mayer, 2005b). 12

Figure 2. Example of a conceptual task used in Experiment 1 and 2 (original version was in German). 19

Figure 3. Example of a causal task used in Experiment 1 and 2 (original version was in German). 21

Figure 4. Example of a procedural task used in Experiment 1 and 2 (original version was in German). 23

Figure 5. Example of a procedural task used in Experiment 3 (original version was in Dutch; obtained from Henny, 2006). 24

Figure 6. Example of pictorial verification questions (a and b) and an integration verification question (c). 39

Figure 7. The difference in means between the multimedia and the text-only condition for free recall as a function of information source and task type. 43

Figure 8. The difference in means between the multimedia and the text-only condition for recall verification accuracy as a function of information source and task type. 45

Figure 9. Means and standard deviations for the text-only and multimedia conditions for performance accuracy as a function of dual task. 70

Figure 10. Means and standard deviations for the text-only and multimedia conditions for performance accuracy as a function of time of testing. 71

Figure 11. Means and standard deviations for the text-only and multimedia condition for description accuracy as a function of dual task and time of testing. 72

Figure 12. Means and standard deviations for the with dual task and without dual task conditions for description accuracy as a function of presentation format and time of testing. 72

Figure 13. CTML adapted to include the process of imagery. 89

10 List of Tables

Table 1 9

Effect Sizes for the Multimedia Effect in Previous Research

Table 2 47

Means and Standard Deviations for Free Recall, Recall Verification Accuracy, and Transfer Verification Accuracy as a Function of Picture-presentation condition, Information Source, and Task Type

Table 3 49

Overview of Significance of the Multimedia Effect and the Effect Sizes (η_{p2}) as a Function of Task Type and Information Source

Table 4 60

Means and Standard Deviations for Study Times, Free Recall, Recall Verification Accuracy, Free Transfer, Transfer Verification Accuracy, and Integration Verification Accuracy as a Function of Picture-presentation condition, Information Source, and Task Type

Table 5 73

Means and Standard Deviations for Performance Accuracy, Description Accuracy as a Function of Dual Task, Presentation Format, and Time of Testing