Helping students remember:
catalytic knowledge and
knowledge outlines with
visual mnemonics.

Erik Sundell

Supervisor: Stina Hallsén

Examiner: Pia-Maria Ivarsson
Abstract
To recall educational content from a lecture or textbook is an efficient way to learn (Karpicke & Blunt, 2011), which is referred to as retrieval practice (Karpicke & Roediger, 2008). It is currently seldom used among students (Karpicke, Butler, & Roediger, 2009), even though it provides benefits such as reducing test anxiety (Agarwal, D’Antonio, Roediger III, McDermott, & McDaniel, 2014), longer lasting memories (Karpicke & Blunt, 2011), and also benefit future learning (Pastötter & Bäuml, 2014). But, in order for retrieval practice to work efficiently, the students must not fail to recall too much of the educational content (Kornell, Bjork, & Garcia, 2011). So in order to help students use retrieval practice, I suggest they are provided with an outlining of the educational content, as this probably helps them remember and recall more of it. In this thesis, I conclude with an experimental approach that it is possible to help students remember such knowledge outlines, and how it can be done. Furthermore, since knowledge such as knowledge about the human anatomy, can be catalytic in the sense that it can enhance future learning (Hattie, 2009; Van Overschelde & Healy, 2001), I also suggest that catalytic knowledge should be identified and made memorable by educators using similar techniques as in this study.

Keywords: experiment, mnemonics, retrieval practice, scaffolding, catalytic knowledge.
# Table of content

Abstract .......................................................................................................................... 2

Introduction ................................................................................................................. 6

1 Theoretical foundation ............................................................................................... 9
   1.1 Previous research .............................................................................................. 9
      1.1.1 Memory retrieval ...................................................................................... 9
      1.1.2 Mnemonic methods ................................................................................. 10
   1.2 A research gap ................................................................................................. 11
   1.3 Theoretical perspectives .................................................................................. 11
      1.3.1 Three memory principles ...................................................................... 12
      1.3.2 The concept of chunking ....................................................................... 13
      1.3.3 The learning style myth ......................................................................... 13

2 Purpose and goal ...................................................................................................... 15
   2.1 Concrete example ............................................................................................ 15
   2.2 The purpose and goal ..................................................................................... 16
   2.3 Experimental hypotheses ............................................................................... 16
      2.3.1 Hypothesis 1 ......................................................................................... 16
      2.3.2 Hypothesis 2 ......................................................................................... 16
      2.3.3 Hypothesis 3 ......................................................................................... 16

3 Method ...................................................................................................................... 17
   3.1 Introduction ...................................................................................................... 17
   3.2 Procedure ......................................................................................................... 17
   3.3 Participants ...................................................................................................... 19
   3.4 Design .............................................................................................................. 20
      3.4.1 Control vs. Treatment ............................................................................. 20
      3.4.2 Picture vs. Sketch vs. Blank ................................................................. 20
      3.4.3 Link vs. Pair ......................................................................................... 21
      3.4.4 Eight groups ......................................................................................... 21
   3.5 Stimuli .............................................................................................................. 21
   3.6 Materials .......................................................................................................... 24
   3.7 Scoring ............................................................................................................ 24
4 Results ........................................................................................................................................ 25
  4.1 Descriptive statistics ........................................................................................................ 25
  4.2 Mean test scores ............................................................................................................... 26
  4.3 Memory durability .............................................................................................................. 28
  4.4 Evaluation of hypotheses ............................................................................................... 30
    4.4.1 Hypothesis 1 ............................................................................................................... 30
    4.4.2 Hypothesis 2 ............................................................................................................... 30
    4.4.3 Hypothesis 3 ............................................................................................................... 30
5 Discussion ................................................................................................................................ 31
  5.1 Results summary ............................................................................................................... 31
  5.2 Experiment issues ............................................................................................................. 31
    5.2.1 Excluded data .............................................................................................................. 31
    5.2.2 Memorization duration ............................................................................................. 32
    5.2.3 Validity ....................................................................................................................... 32
  5.3 Explaining results .............................................................................................................. 33
    5.3.1 Number of associations ............................................................................................. 34
    5.3.2 Association quality ..................................................................................................... 34
    5.3.3 Memory maintenance and repairability ..................................................................... 35
    5.3.4 Distractions ............................................................................................................... 37
    5.3.5 Summary .................................................................................................................... 37
  5.4 Applications and future research ..................................................................................... 38
    5.4.1 Knowledge outlines ................................................................................................. 38
    5.4.2 Factual foundations ................................................................................................. 40
  5.5 Conclusions ....................................................................................................................... 41
References .................................................................................................................................... 43
Image references ....................................................................................................................... 45
Appendix A – Additional results ............................................................................................. 46
  A.1 Experiment durations ...................................................................................................... 46
  A.2 Test score details ............................................................................................................. 47
    A.2.1 Test score distributions ............................................................................................ 47
    A.2.2 Test acers .................................................................................................................. 47
A.3 Questionnaire data

A.3.1 Test score estimations

A.3.2 Memorization visualization

A.3.3 Self-recall attempts

Appendix B – Experience mnemonic methods

B.1 The Russian doll method

B.2 The method of loci

B.3 Thesis outline

Appendix C – A detailed experiment overview

Appendix D – Participant interactions

Appendix E – Experiment stimuli

Appendix F – Experiment administration
Introduction

If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows.

– David Ausubel (1968, p. iv)

Having previous knowledge makes related knowledge more interesting. It has been found that smaller gaps in our previous knowledge can motivate us to fill in the gap, but we are discouraged by too large gaps as we then have too little knowledge to relate to (Hattie & Yates, 2014, p. 23). It has also been shown that knowledge be catalytic, that knowledge can act as a catalyst, making it easier to learn more (Hattie, 2009). Even basic facts can be catalytic (Van Overschelde & Healy, 2001; Willingham, 2010, p. 32).

If it took me little-enough time, there are several things I would make sure to learn. For example, all countries of the world, their locations on the world map, their capitals and their national flags. Would you also consider learning that if it took you little enough time? Do you agree that it is relevant to consider the time required to learn something? What if a chemistry teacher could help students memorize the periodic table in an hour? What if a computer program could help teachers to learn the first name of all students at the school quickly? What if teachers could help students remember an outline of a lecture’s content, a knowledge outline, in one minute?

It is my experience that everyone can memorize information efficiently, if provided proper support or effective supportive material that I refer to as scaffolding. With this study I aim to enable educators to help their students remember catalytic knowledge, especially knowledge outlines, by using visual mnemonics.
The word *mnemonic* (the first m is silent) is derived from a Greek word that means memory. A mnemonic is something that makes some information more memorable, such as the mnemonic “Eddie Ate Dynamite Good Bye Eddie”, which can make the notes of the guitar’s six strings, EADGBE, more memorable. In this mnemonic, letters were made into a sentence to make the letters more memorable. Methods to make things more memorable, such as making letters into amusing sentences, are called *mnemonic methods*.

To conclude, one can say that mnemonic methods work by associating information to something that is more memorable. Memory athletes, competing in the ability to remember new information, all seem to agree that mnemonic methods that associate information to something visual, *visual mnemonic methods*, are the most efficient mnemonic methods. They also seem to agree on that everybody can learn to utilize such methods (Foer, 2011; Ribbing, 2013; von Essen, 2014a; Zogaj, 2012). Their belief is supported by neuroscience, which concludes that memory athletes ability doesn’t arise from an exceptional intellectual ability or structural brain differences, but from using a visualization strategy to remember (Maguire, Valentine, Wilding, & Kapur, 2003).

I will utilize visual mnemonics to see if it is possible to help students memorize an outline of a lecture or textbook, a *knowledge outline*, and I will argue that providing mnemonics to help students remember knowledge outlines will synergize well with a study method called *retrieval practice*. It is sadly a relatively unknown and unused way to study (Karpicke et al., 2009), even though it has been shown to be effective (Butler, 2010; Carpenter, 2012; Karpicke & Roediger, 2008), and give benefits such as reducing test anxiety (Agarwal et al., 2014).

Three Swedish memory athletes have inspired me to do this study, Mattias Ribbing, Jonas von Essen and Idriz Zogaj. They all believe that mnemonics and mnemonic methods can
be used to empower education, and are all making effort for this to happen. As a teacher student, I believe I might miss out on very useful knowledge if I do not consider what memory athletes have learned. If you want some experience with visual mnemonics right now, go to Appendix B.
1 Theoretical foundation

1.1 Previous research

Essential for the purpose of this study, is research about the study method referred to as retrieval practice. I will start there, and then introduce some research about mnemonic methods.

**Memory retrieval.** To retrieve a memory is to recall it, to make it vivid again. Try to visualize a banana for example. By doing that, you retrieved several details about it from your memory. Now, keep reading but avoid thinking of an apple, its shape, color or stem. Did you manage to avoid retrieving any memories of an apple?

To retrieve a memory, one commonly make use of retrieval cues, often automatically without thinking and even if you did not want to. The words about the apple were acting as retrieval cues to remember details about apples. A retrieval cue is something that stimulates the retrieval of a memory. It can for example be a smell from the kitchen that stimulates you to remember what you previously put inside the oven, or a certain song that causes you to suddenly think of a specific moment a long time ago.

It has been concluded that a lack of retrieval cues often is the reason for forgetting (Willingham, 2008). Similarly, it has also been found that cued recall is easier than free recall, which is to say that it is easier to remember when you are provided with retrieval cues than without them (Goldstein, 2011, p. 182).

*Retrieval practice* is a study method where the student focuses on retrieving memories. To recall or in other words retrieve memories, is often a part of tasks given to students, but it seldom is the task itself. Even though retrieval practice is efficient (Butler, 2010; Carpenter, 2012; Karpicke & Roediger, 2008), it is seldom used by itself (Karpicke et al., 2009). It has been shown to improve actual classroom learning (Agarwal, Bain, & Chamberlain, 2012), reduce test
anxiety (Agarwal et al., 2014), improve long term retention (Karpicke & Blunt, 2011) and lead to improved future learning (Pastötter & Bäuml, 2014). Karpicke and Blunt convey a belief that retrieval play a vital role in learning, based on research in cognitive science (2011, p. 772). They say that “Retrieval is not merely a readout of the knowledge stored in one’s mind; the act of reconstructing knowledge itself enhances learning.” (2011, p. 774).

Retrieval practice is based on the testing effect, that putting your memory to the test and forcing yourself to retrieve what you remember, for example makes the successfully retrieved memories easier to retrieve in the future (Roediger & Karpicke, 2006). This means that tests that are too hard, where too much of the recall attempts fail, are inefficient (Kornell et al., 2011).

While retrieval practice is efficient for college students, it has been found that elementary school students need more guidance while performing it (Karpicke, Blunt, Smith, & Karpicke, 2014).

**Mnemonic methods.** Research about mnemonic methods exists, also within educational contexts, such as for foreign vocabulary learning. The keyword method for example helps the user remember associations, such as between a word and its meaning (Pressley, Levin, & Delaney, 1982). It is done by connecting two visual retrieval cues by visualizing an interaction between them, a process that make the association very memorable. For example, to learn the Swedish word for ice cream is spelled “glass”, you could visualize an ice cream, squished to a glass window. The keyword method, as well as retrieval practice, has been found able to empower learning of new words in a foreign language (Fritz, Morris, Acton, Voelkel, & Etkind, 2007).

Other mnemonic methods, such as the method of loci, has been shown to efficiently help the user remember a sequence of information (Roediger, 1980). To use the method of loci, you
encode information to visual retrieval cues, and make them memorable by visualizing them to interact with a part of something you visually know very well, for example your kitchen, your bicycle or your own body. For example, to remember a shopping list, you can imagine a leaking flour package on your head, bananas sticking out of your ears and an egg in your left eye socket.

Note that the visualizations doesn’t need to be bizarre for these mnemonic methods to work (Bower, 1970). Some guidelines from memory athletes are that the visualizations should be made large, three-dimensional, in color and with details (Kozarenko, 2006, p. 78; Ribbing, 2013, p. 91).

1.2 A research gap

Mnemonics and retrieval practice has separately been shown able to be beneficial in educational contexts. Some research have even been done when both mnemonics and retrieval practice is used in parallel (Fritz et al., 2007). However, to my knowledge, no research has been made that evaluates if mnemonics can be utilized specifically to enhance retrieval practice in general.

With this study, I hope to conclude if and how visual mnemonics could help students remember knowledge outlines. It is my hope students using such outlines to guide retrieval practice will make them successfully recall more of the content, which is important for the benefits of retrieval practice to show (Kornell et al., 2011). I will present a more concrete picture in the Purpose and goal section of this text.

1.3 Theoretical perspectives

In this section, I will present three principles of memory, the concept of chunking, and research about learning styles.
Three memory principles. Daniel Willingham, Professor of Psychology at the University of Virginia with a Ph.D. in cognitive psychology, has stated three principles of memory that I will relate results to (Willingham, 2008).

The first principle is that *memory is a residue of thought*. For example, it has been concluded that you need to pay some attention to what your ears pick up if you mean to remember it (Goldstein, 2011, pp. 83–87). The brain must process a stimuli to make it memorable (Hattie & Yates, 2014, Chapter 13).

The second principle is that *memories are lost mostly due to missing or ambiguous cues*. Have you experienced the tip-of-the-tongue phenomena? That is when you feel confident on knowing something, but you cannot recall it. It is a good example of the principle, how a lack of retrieval cues often leads to inability to remember. This principle also finds support in research that concludes that it is easier to recognize, than to freely recall something, as well as one tend to quickly forget isolated facts (Hattie & Yates, 2014, Chapter 13). It is also supported by Tulving and Pearlstone’s experiment, which found that *retrieval cues* is a great aid to the memory (Tulving & Pearlstone, 1966; Goldstein, 2011, pp. 182–183).

The final principle is that *individuals’ assessments of their own knowledge are fallible* (Willingham, 2008). In other words, we are often not so good at evaluating how well we know something. Rereading is currently the predominant study practice, at least among a set of college students in USA (Karpicke et al., 2009), and I believe rereading often can become a process of recognition rather than actually processing the content. However, when you are tested, you probably need to recall it freely rather than recognize it, and it will be harder. Maybe this is why we often do not assess our knowledge so well? Because we have not yet fully tested it, only become familiar with it, and become able to recognize it, which is easier. And maybe that is why
retrieval practice have been found to reduce test anxiety (Agarwal et al., 2014), because one is forced to test one's capability of recalling it freely.

**The concept of chunking.** A concept of relevance for this text is chunking. A chunk has been defined as “[…] a collection of elements having strong associations with one another, but weak associations with elements within other chunks” (Gobet et al., 2001, p. 236). Chunking is about creating internal associations between a set of memory elements. The more associations your brain forms to a memory, the more ways the brain gets to retrieve them. Therefore, by chunking, you create an increasing amount of fail-safes to forgetting, as Willingham’s second principle states that missing cues or associations often is the reason for forgetting. One tends to forget isolated facts very quickly for example (Hattie & Yates, 2014, Chapter 13). To chunk is to make something very connected or associated, the opposite of making something isolated. It makes the content more memorable.

For example, it is probably harder for us to remember the letters within the word “mneietepxr” than the letters in the word “experiment”, even though they are the same, since we probably recognize the latter as a chunk, but not the former. Another example is that chess grandmasters can recognize many situations in chess games well, and due to this, are able to remember played chess games better than amateurs are.

**The learning styles myth.** It is commonly said that everybody learns differently, but it is a very broad statement that when specified more clearly could mean many different things. Some of these things are not true. For example, the claim that some learn better visually and others better auditorily, or in researchers’ terminology, that we have different *learning styles*. Extensive research has simply not found evidence for some of us to learn better visually, auditorily or kinesthetically, just because it matches our preferred learning style. Also, no educational benefits
have been found for utilizing knowledge about someone’s learning style (Pashler, McDaniel, Rohrer, & Bjork, 2008).

Notice that this research does not contradict the memory athletes’ claim that visual mnemonic methods are more efficient than other mnemonic methods (Introduction, p. 7), because they do not claim that it will differ between persons while the loose speculations of learning styles does.
2 Purpose and goal

“Retrieval is not merely a readout of the knowledge stored in one’s mind; the act of reconstructing knowledge itself enhances learning.”

– Jeffrey D. Karpicke and Janell R. Blunt (2011, p. 774)

2.1 Concrete example

Let me attempt to provide you something specific that you can relate the purpose and goal of this study to. Let me talk about ancient Greek speakers.

Some speakers among the ancient Greeks used a mnemonic method to memorize sequences of images that reminded them of speeches’ key points. In other words, they memorized outlines of their speeches. Memorizing outlines like this allowed them to give the speeches as they intended without forgetting something important.¹ I speculate that we often use outlines to remember better without noticing it. For example if you forgot your keys, do you sometimes think backwards of the places you have been to recently? Then you follow an outline.

Now let us consider today’s educational system. What would happen if students made internal speeches for themselves, with their mind’s inner voice, about educational content presented by a lecturer? Would they learn something by giving such internal speeches? I believe so, because it would be retrieval practice and very active processing of the content. However, I also speculate that memorizing an outline of the speech as the ancient Greek speakers did might help them. This study will investigate if educators can help students remember such outlines, in the hope that future research will demonstrate that it would be beneficial for students learning if they use them to give internal speeches.

¹ This is based on my basic understanding and interpretation of history. If you are curious to investigate yourself, I suspect “The Art of Memory” by Frances A. Yates published 1966 might be a good read.
2.2 The purpose and goal

The primary purpose of this study is to act as a foundation for future research about the uses of mnemonics in education. The study’s goal is to find a way to scaffold memorization of an outline of some educational content, represented as a sequence of visual retrieval cues, such that free recall of the outline 24 hours after the memorization is improved compared to a self-directed memorization.

2.3 Experimental hypotheses

I made three hypotheses. After reading the method section, you will be able to understand the more precise formulations of the hypotheses provided in the footnotes.

Hypothesis 1 – Mean test scores. I predict that the participants receiving mnemonic scaffolding will successfully recall more objects from a memorization procedure, just briefly after it as well as about 24 hours later.²

Hypothesis 2 – Memory durability. I predict that the participants receiving mnemonic scaffolding will form memories that are more durable.³

Hypothesis 3 – Superiority. I predict that the participants receiving mnemonic scaffolding with the most sophisticated provided imagery, will recall the most objects overall in the free recall tests and form the most durable memories.⁴

² I predict that the treatment groups will receive a significantly higher mean test score in a) the initial and b) the delayed free recall test than the control groups.

³ I predict that the fraction of participants lowering their score in between the two tests will be significantly lower for the treatment groups than for the control groups.

⁴ I predict that the treatment group Link Picture will a) receive the highest mean test scores among the treatment groups, and b) have the lowest fraction of participants lowering their score in between the two tests.
3 Method

3.1 Introduction

I created an experiment to evaluate how well variations of mnemonic scaffolding actually worked for participants in treatment groups, and compared it to how well participants performed without such scaffolding in control groups. Let me explain why I decided to use the mnemonic method that I used to guide the creation of the mnemonic scaffolding, consisting of instructions and imagery, which was provided the participants in the treatment groups.

The mnemonic method I utilized was the Russian doll method (RDM), and I mainly choose it because one can get started with it without preparation. It did for example not require references to some visual background as the method of loci does. The RDM is very similar to the chain method, also called the link method (Kozarenko, 2006, pp. 141–142). Both these methods work by visualizing objects pairwise to form links that together form a chain, as demonstrated in Table 1.

<table>
<thead>
<tr>
<th>Link</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>Russian doll</td>
</tr>
<tr>
<td>BC</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Visualization examples from the Russian doll method and the chain method. Please refer to the references for image attribution.

Note: Either the following object is made relatively very small (RDM), or positioned to the right (chain/link method).

I made the choice to use the RDM over the chain method by guessing that the RDM’s imagery would be more powerful because it systematically provides a noticeable detail to peak interest and focus attention on initially, namely the object depicted as smaller.

3.2 Procedure

The experiment was designed to test if instructions of a mnemonic method, with various scaffolding imagery, would aid the memorization of visualizable objects, such as a table tennis
racket. The idea of the experimental design was to assign participants randomly to different
groups, and then to give the groups the common task of memorizing twelve objects, and compare
their performance based on the differences introduced between the groups. For example, the
treatment groups received mnemonic instructions while control groups did not.

Let me elaborate on the basic experiment overview in Figure 1. Participants were initially were randomly assigned to experimental groups, called treatment groups if they received mnemonic instructions and control groups if they did not. They then spent up to a maximum of four minutes to memorize twelve objects. After the memorization, they were asked some questions, to thereafter do the initial free recall test, where they demonstrated how many of the twelve objects they could recall. After 24 hours, an email reminder was sent out asking the participant to come back and repeat the test.

Figure 1. A basic overview of the experiment.
3.3 Participants

I recruited voluntary participants for the experiment using social media. I did not use an economic incentive, but I framed a report with data visualization on their performance as an incentive to participate and to return 24 hours later for the delayed test.

All participants had at least a basic knowledge of Swedish. The mean age among the participants was 35 years ($N = 321$, $SD = 11$ years). Nearly twice as many females (64%) as males (36%) participated ($N = 321$). I also suspect that many of the recruited participants were a) university students and b) teachers in the Swedish school system, because of my personal network and social marketing of the experiment within Facebook.

I filtered the collected data by certain requirements, presented in Table 2 below.

**Table 2.** Participant information. The table shows how many participants started out ($N = 665$) and how many entries ended up contributing to usable data ($N = 321$). All numbers represent a participant count.

<table>
<thead>
<tr>
<th>Group</th>
<th>Data exclusion and dropout reasons</th>
<th>Participant entry summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Corrupt entries</td>
<td>Aborted early</td>
</tr>
<tr>
<td>Link Picture</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Link Sketch</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>Pair Picture</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>Pair Sketch</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>Link/Pair Blank</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Control Picture</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Control Sketch</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Control Blank</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>

**Summary**

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>83</td>
<td>665</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>344</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>321</td>
</tr>
</tbody>
</table>

Note: Corrupt entries represent participants that experienced technical errors, which were identified by automated error reports. Aborted early represent participants closing the webpage before the initial part was completed. Skipped memorization represent participants that spent less than 30 seconds on the memorization or less than two seconds per association in the treatment groups. Never returned represent participants that never returned for a delayed recall test. Also note that participants belonging to multiple categories, only is listed in the left most category.
3.4 Design

In this section, I will present the details differentiating the eight experiment groups. Table 3 below summarizes several parts of this section.

**Table 3.** Experimental design. A structural overview of the groups.

<table>
<thead>
<tr>
<th>Group type</th>
<th>Treatment variation</th>
<th>Scaffolding material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Link</td>
<td>Link Picture</td>
</tr>
<tr>
<td></td>
<td>Pair</td>
<td>Pair Picture</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control Picture</td>
</tr>
<tr>
<td>Control</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Link Sketch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pair Sketch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control Sketch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Link/Pair Blank</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control Blank</td>
</tr>
</tbody>
</table>

**Control vs. Treatment.** The differences between the three control and the five treatment groups related to the memorization instructions as well as the memorization procedure.

The memorization instructions for the control groups said that they should attempt to memorize to the best of their ability, a self-directed memorization. The treatment groups were instructed to use the *Russian doll method* that they got to read about briefly (Appendix E, Screenshot 5a-b).

The memorization procedure differed in the sense that control groups would see the same view during the memorization, containing all twelve objects, for up to four minutes. The treatment groups on the other hand, would automatically cycle through twelve different views, each containing two objects. The twelve views formed a closed cycle, like the pairwise cycle of the four letters AB, BC, CD and DA. To represent the various sizes of the objects in each view the cycle could be exemplified by Ab, Bc, Cd and Da. Each view was shown for up to 20 seconds, which sums up to four minutes. Table 4 on page 23 show some of the cycled views.

**Picture vs. Sketch vs. Blank.** During the memorization procedure, all groups were able to see the names of the objects written out, but some received additional imagery. Some groups were provided with full color photographic pictures of the objects. Other groups were provided
with black and white sketches of the objects. The final category *Blank*, represent groups that were not provided any additional imagery.

**Link vs. Pair.** Among the five treatment groups, two were classified as *Link* groups and two as *Pair* groups. *Pair* groups were shown *two separate pictures of objects*, while *Link* groups were provided with *one picture containing both objects visually connected*, depicting how they had been instructed to connect the objects by visualization.

**Eight groups.** Let me summarize. There were three control groups, internally distinguished with *Picture, Sketch* and *Blank* labels, and five treatment groups. The treatment groups had two *Link* groups and two *Pair* groups, internally distinguished with *Picture* and *Sketch* labels. The final treatment group that was not provided additional imagery, could not due to this be classified as either *Link* or *Pair*, and was due to this named *Link/Pair Blank*. This grouping allowed me to identify effects of provided imagery somewhat separately from the instructions.

### 3.5 Stimuli

Two different memorization instructions were created, one for the three control groups and one for the five treatment groups. The control groups were asked to memorize to the best of their ability (Appendix E, Screenshot 5a), while the treatment groups were asked to use the Russian doll method (RDM) which was introduced in the method-section’s introduction on page 17 (Appendix E, Screenshot 5b; Kozarenko, 2006, p. 142).

In brief, to use the RDM, one should visualize objects pairwise to create links that together form a chain. One object is visualized as bigger than the other, which is visualized as very small, and attached to the big object. To connect more than two objects, one should zoom in on the previously small object and make it bigger in one’s *visuospatial sketchpad*, one’s inner
eye. One would then attach another small object to the big object like before, etc. See the pictures that *Link Picture* group were provided during the memorization procedure in Table 4 for an example of what you are supposed to visualize.

When visualizing, four things are considered important by experienced practitioners in order to form rigid memories. One should aim to visualize: large, three-dimensionally, in color and with details (Kozarenko, 2006, p. 78; Ribbing, 2013, p. 91). Although, to make the treatment groups’ instruction briefer, this was not explicitly written out.

The table on the next page demonstrates the varying imagery provided the experiment groups.
Table 4. The experiments memorization procedure, depicted for the eight experiment groups.

<table>
<thead>
<tr>
<th>Depicted moment</th>
<th>Treatment groups</th>
<th>Control groups</th>
<th>Control Blank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Link Picture</td>
<td>Link Sketch</td>
<td>Pair Picture</td>
</tr>
<tr>
<td>First 20 seconds</td>
<td>Link Picture</td>
<td>Link Sketch</td>
<td>Pair Picture</td>
</tr>
<tr>
<td>Following 20 seconds</td>
<td>Link Picture</td>
<td>Link Sketch</td>
<td>Pair Picture</td>
</tr>
<tr>
<td>Following 20 seconds</td>
<td>Link Picture</td>
<td>Link Sketch</td>
<td>Pair Picture</td>
</tr>
<tr>
<td></td>
<td>Control Picture</td>
<td>Control Sketch</td>
<td>Control Blank</td>
</tr>
<tr>
<td>The full duration of four minutes.</td>
<td>Control Picture</td>
<td>Control Sketch</td>
<td>Control Blank</td>
</tr>
</tbody>
</table>

*Note: Please see the image reference section for image attribution.*
3.6 Materials

For the purpose of this experiment, I built a webpage, which contained the experiment, and some tools to help me administer it.\(^5\)

The administration tools was made to; automatically remind participants with emails, automatically generate the participants result reports and email them, facilitate a manual validation of the unique test answers, automatically grade tests and display error logs from the experiment webpage as well as from the administration tool itself. It was also made to notify me directly about any critical errors using SMS messages. Screenshots from the administration software is available in Appendix F.

3.7 Scoring

About three hundred unique answers from the free recall tests were collected, due to massive controversies about how to spell the recalled objects (Appendix F, Screenshot 10). All of the unique answers were classified as either wrong or correctly representing a specific object. To avoid any bias while manually classifying the answers, the classification was done insulated from any information connecting the answers to other participant data.

\(^5\) Three major parts worked together, a) the front end, b) the back end and c) the administration. The front end, which represents what the participants got to see and the code running in their web-browsers, was built using HTML, CSS and JavaScript and utilized the front-end framework EmberJS (.com) as base to build on. The back end, which represents the part that the front end communicates with, was a JSON document store hosted by Firebase (.com) wherein several security rules were setup to ensure only authorized access of the data. Administration tools and utilities were created using a Google Spreadsheet in combination with Google Apps Scripts and an emailing service provided by Mandrill (.com).
4 Results

The experiment tested how well participants of eight experimental groups had memorized twelve objects by asking them to write down the objects they remembered, about one minute as well as 24 hours after a memorization procedure. This section will describe how well the participants of the groups performed during the two first test, which will be referred to as the initial test and the delayed test. Additional results are found in Appendix A.

In the end of the results section, I will summarize evidence relating to the hypotheses.

4.1 Descriptive statistics

Instead of writing out the large amount of data for you to grasp by your own, I will attempt to summarize it using for example; the number of participants ($N$), the number of participants in one specific group or subset ($n$), the mean value ($M$), and the median ($Md$), and the standard deviation ($SD$). The mean value is the average. The median is the value that would be in the center if all values were sorted in ascending or descending order. The standard deviation indicates how far off a specific observation is from the mean value.

Let me also describe, that with an increasing amount of data ($N$ or $n$), calculated mean values ($M$) become more accurate. With too little data, they might be far off the true mean value, which is what you would get with an infinite amount of data. In this text, a confidence interval with a 95% confidence level is therefore provided with the mean values. They represent a range where the true mean value with a 95% confidence will be found within. It can look like this.

You and your friends received the highest mean test score ($M = 11.2 [10.8, 11.6]$).
4.2 Mean test scores

After the memorization, the participants were asked some background questions that took about one minute to answer. After that, they did their first recall test. They got one point for each different object they could write down from the memorization of the twelve objects.

The mean test scores for the eight experiment groups are shown in Figure 3a. The five treatment groups are seen to the left, and the three control groups to the right. The group with the highest mean test score was Link Picture \((M = 11.2 \ [10.6, \ 11.6])\), and the group with the lowest mean test score was Control Blank \((M = 10.5 \ [9.8, \ 11.2])\). Link Picture was the group that was provided with the most sophisticated imagery during the memorization procedure, and it was my belief that it would receive the highest mean test score, mainly because I believed that the imagery provided Link Picture would make everyone able to fully understand the treatment groups’ memorization instructions.

To analyze this data further, consider if we would compare eight different persons, who had tossed a coin a hundred times, if we counted how many times they got tails up, we would probably find that some persons received more tails than the others, even though no actual differences ought to be concluded from differences seen in such experiment. In such situation, an analysis of variance (ANOVA) could be useful.

An ANOVA tells us how great the risk is, that the observed differences could have arisen from random variations, rather than that a person actually consistently tossed either more or less...
tails than the others. The p-value represents this risk. The amount of risk that the researcher decides to be tolerated ahead of time is called the significance level, often denoted with $\alpha$. I have decided to use a significance level of .05. An ANOVA’s reported p-value would only be lower than .05, in 5% of the situations if the results were actually only a consequence of random variations, such as when tossing coins high enough.

An ANOVA on the groups’ mean test scores in Figure 3a resulted in a high p-value far above .05, $F(7, 313) = 0.74, p = .647$. That means not enough evidence was found to assume that the differences between groups in the initial test was due to something else than random variations. I state this because I require the p-value ($p = .647$), now interpreted as the risk of drawing a faulty conclusion, to be less than the significance level ($\alpha = .05$). An ANOVA between the treatment groups and the control groups was also found to be insignificant, $F(1, 319) = 0.43$, $p = .511$. We have thus so far not seen enough evidence to conclude that the treatment had any benefit. However, this test occurred very quickly after the memorization, so let us look at the results of the recall test after 24 hours.

In the delayed test after 24 hours, the mean test scores were lower overall, and the differences in the mean test score between groups became large enough for an ANOVA to be significant, $F(7, 313) = 2.5, p = .017$. In other words, enough evidence was found to assume that the differences shown between groups in their mean delayed test score, were to some degree due to the differences between the groups, and not only due to random variations.
In the delayed test, the treatment groups \((M = 9.9 \ [9.6, 10.2])\) received a higher mean score than the control groups \((M = 9.0 \ [8.5, 9.5])\).\(^6\)

Among the groups in the delayed test, the treatment group *Link Picture* again received the highest mean test score \((M = 10.5 \ [9.8, 11.1])\), and the group with second highest score \((M = 10.0 \ [9.1, 10.8])\) was the treatment group *Link/Pair Blank*, which was the treatment group that wasn’t provided images along the object names during the memorization. The strongest control group, *Control Sketch* \((M = 9.5 \ [8.6, 10.4])\), performed similar to the weakest treatment group, *Pair Picture* \((M = 9.5 \ [8.7, 10.2])\).

To summarize this heading, I want to point out that differences between groups were confirmed in the delayed test, and *Link Picture* received the highest mean test score, with *Link/Pair Blank* as second highest. Also, note that providing sophisticated imagery as in *Link Picture* or no imagery at all as in *Link/Pair Blank*, both seems like reasonable choices when scaffolding memorization using the RDM.

### 4.3 Memory durability

This section focuses on presenting data to indicate how durable were the participants’ memories were. How much did they forget in between the two tests?

Most participants let about 24 hours pass in the forced delay between the tests. During this time, 53% of the participants forgot one or more objects \((N = 321, 95\% \ CI [48\%, 59\%])\). This section is about the question if some participants had formed memories that were more durable, and had become less likely to forget.

I noticed that the treatment groups forgot significantly less than the control groups, \(F(1, 319) = 9.5, p = .002\), which confirms Hypothesis 2.

\(^6\) An ANOVA comparing the treatment groups with the control groups were significant, \(F(1, 319) = 10.1, p = .002\).
I also noticed participants that aced the initial test, got the highest score on it, and seldom forgot objects during the forced delay between the tests. Only 37% of the participants that aced the initial test \((n = 157, 95\% \text{ CI} [29\%, 44\%])\) lowered their score in the delayed test, while 69% of the other participants \((n = 164, 95\% \text{ CI} [62\%, 76\%])\) lowered their score.

Finally, I noticed that participants in the Link Picture and Link/Pair Blank seemed to have formed especially rigid memories, because very few of them forgot any object if they had initially recalled every object successfully during the initial test. Link Picture and Link/Pair Blank’s initial test acers lowered their score only 15% of the time, which can compare to 41% for Control Picture or 63% for Control Blank.

The participants that did not ace the initial test were overall likely to lower their score in between the tests.

To summarize this heading about memory durability, I will point out that the treatment groups overall, as well as the participants that aced the initial test, formed memories that were more durable. The participants within Link Picture and Link/Pair Blank that aced the initial test formed the most durable memories.
4.4 Evaluation of hypotheses

**Hypothesis 1 – Mean test scores.** I predicted that the treatment groups’ participants would receive a significantly higher mean test score a) in the initial test and b) in the delayed test, than the control groups. By an analysis of the variance, part a) of this hypothesis was rejected, $F(1, 319) = 0.4, p = .511$, and part b) was accepted, $F(1, 319) = 10.1, p = .002$.

**Hypothesis 2 – Memory durability.** I predicted that the fraction of participants lowering their score in between the two tests would be significantly lower for the treatment groups. An analysis of variance confirmed this hypothesis, $F(1, 319) = 9.5, p = .002$.

**Hypothesis 3 – Superiority.** I predicted that the treatment group Link Picture would a) receive the highest mean test scores among the treatment groups, and b) have the lowest fraction of participants lowering their score in between the two tests. All parts of the hypothesis were confirmed. The data underlying the conclusion is presented in the table below.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Mean test scores in...</th>
<th>Score degradation among...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Initial test</td>
<td>Delayed test</td>
</tr>
<tr>
<td>Link Picture</td>
<td>42</td>
<td>11.2</td>
<td>10.5</td>
</tr>
<tr>
<td>Link Sketch</td>
<td>43</td>
<td>10.9</td>
<td>9.8</td>
</tr>
<tr>
<td>Pair Picture</td>
<td>38</td>
<td>10.6</td>
<td>9.5</td>
</tr>
<tr>
<td>Pair Sketch</td>
<td>38</td>
<td>10.6</td>
<td>9.7</td>
</tr>
<tr>
<td>Link/Pair Blank</td>
<td>33</td>
<td>10.6</td>
<td>10.0</td>
</tr>
<tr>
<td>Control Picture</td>
<td>44</td>
<td>10.8</td>
<td>9.0</td>
</tr>
<tr>
<td>Control Sketch</td>
<td>40</td>
<td>10.8</td>
<td>9.5</td>
</tr>
<tr>
<td>Control Blank</td>
<td>43</td>
<td>10.5</td>
<td>8.5</td>
</tr>
</tbody>
</table>

*Note:* Together *Figure 3a-b* and *Figure 4a-b* represent all data in the table except the total score degradation, which was relevant for Hypothesis 3.
5 Discussion

In this section, I summarize and discuss the results, highlight potential issues, and elaborate on practical applications of the results in an educational context.

5.1 Results summary

The treatment groups received a higher mean test score and formed more durable memories than the control groups. The group Link Picture performed better than the other groups overall. That group had: the highest mean test scores in both test (Results, Figure 3a-b), the highest fraction of participants that aced both tests (Appendix A, Figure 6a-b), and together with Link/Pair Blank it had the lowest fraction of participants that degraded an initially perfect score during the 24 hours delay between the tests (Results, Figure 4a). The initial acers of Link Picture, as well as Link/Pair Blank, had formed especially durable memories compare to the other groups. Taken together this means that all hypotheses except Hypothesis 1a about the initial test were confirmed.

Participants in the treatment groups spent only about 70% of the time spent by the treatment groups’ participants on the memorization procedure (Appendix A, Table 6). This issue will be discussed in the following section.

5.2 Experiment issues

Let me shed some light on issues I have identified with the experimental data I collected, and how that affects what we can conclude from it.

Excluded data. To exclude of one participant’s data entry at random would be as if an experiment group had received one less participant. Since participants were randomly assigned to the groups initially, such random exclusion of data would not cause issues for the analysis. However, if more or less data was excluded depending on the group, issues could arise. For
example, I believe that participants especially in Link/Pair Blank, which had the most dropouts, became disappointed with the memorization procedure, since no imagery was provided. So perhaps a greater ratio of impatient participants cancelled the experiment in this group. If impatient participants overall also tended to receive lower test scores, the mean test score for Link/Pair Blank would be skewed to look better than it should, due to this.

How much problematic data exclusion have occurred is hard to evaluate, but roughly 50% of the participants that initiated the experiment didn’t complete it or was excluded from the analysis for other reasons such as spending unreasonably little time on the memorization of objects. You can look back at Table 2 in the method-section 3.3 for more details.

Memorization duration. Another thing to consider is that the control groups spent only about 70% of the time that the treatment groups spent on the memorization (Appendix A, Table 6). So any differences between a treatment group and a control group’s test performance could at least partially be argued to be caused by the difference in time spent on the memorization.

So why did the control groups spend less time on the task? I speculate that the control groups experienced the task to be less stimulating with their single view that did not change. I also speculate that the treatment groups’ participants spent more time on the memorization because they were given a specific goal to achieve during the memorization procedure, which was to visualize according to the RDM.

Validity. Issues with the data are present, so can conclusions still be drawn? Yes, but the issues must be taken into account. To evaluate the impact of the collected data’s issues, consider the following thoughts.

First, among groups with similar dropout rates and memorization durations, significantly different performance could still be found (Results, Figure 4a), which indicates that other factors
also significantly influence the results. Secondly, even though we cannot conclude that the treatment groups’ memorization was more time efficient than the control groups’ due to the systematic difference in time spent, we can conclude that the treatment gave better result, and that the treatment instructions and/or procedure caused the participants to spend more time on the memorization task.

Now that we are aware of some of the issues with the collected data, let us try to explain the results.

5.3 Explaining results

The differences between treatment and control groups could at least partially be explained by the difference in time spent, but not fully, since significant differences were found among the treatment groups results even though they spent similar amount of time on the memorization. Therefore, I will now try to explain some of the results using other factors as well.

For example, why did Link Picture and Link/Pair Blank’s participants that initially aced, form such durable structures compared to all other groups? I want to avoid giving a too simple explanation, since the answer probably depends on many factors, but I want to highlight four factors that I will argue contribute to the explanation. They are; a) the number of associations made to the memorized objects, b) the quality of these associations, c) the ability to maintain and repair the formed memory structures, and d) distractions during memorization.
**Number of associations.** Let us first consider the amount of associations made to the memorized objects. The treatment groups’ instructions were to visualize the objects pairwise, until they had form a closed chain, as is represented by Memory state 1 to the right, with the simplification of five objects instead of twelve.

Notice that that every object memorized would have two visual associations connecting it to the neighboring objects. Willingham’s second memory principle states that missing cues or associations is a common reason for forgetting (Willingham, 2008). This along with research about chunking, indicate that crafting multiple associations to something, is a way to reduce the risk of forgetting it.

I believe that the participants of the treatment groups memorized the objects with more consciously crafted associations than the participants in the control groups did overall, and that this contributes to the explanation of the observed difference between treatment and control groups. In other words, using the Russian doll method, as the treatment groups’ participants did, might have led to formation of a better-connected memory structure, which could contribute to explaining why their memories were observed to be more durable.

**Association quality.** Now, let us consider how variations in the quality of the associations might contribute to explaining the results.

Associations’ strength can vary. Using associations or reestablishing them will make them stronger, as Willingham’s first memory principle states that memory is a residue of thought (2008). Memory athletes, well versed in the art of remembering, seem to have reached a...
consensus about crafting associations. They claim that crafting visual associations is the most time efficient way to form strong and rigid associations, and that this is true for everyone, not just those that for example have stated they have a visual learning style (Foer, 2011; Zogaj, 2012; von Essen, 2014a; Ribbing, 2013).

During the memorization, the treatment groups’ participants presumably dedicated their time on crafting visual associations. They had been given details on how to make them rigid based on memory athletes’ expertise. Due to this, I believe that the participants of the treatment groups overall created more rigid associations. Let me support this by some data. The participants were asked to evaluate their degree of visualization during the memorization, as well as what score they thought they would receive during the delayed test. One might think that their estimate of their score on the delayed test would be the best indicator of their actual score among these questions, but they were actually equally efficient (Appendix A, p. 48).

Memory maintenance and repairability. Finally, let us consider how a varying ability to maintain and repair memories might contribute to explaining the observed results.

Retrieving memories strengthen them (Karpicke & Roediger, 2008), so retrieving memories can be thought of as maintenance. This idea is supported by data collected by the participants in the experiment. It was found that participants’ reported self-recall attempts, the amount of times they by themselves had tried to recall the memories of the memorized objects, was found to correlate with the delayed test score. Participants with more reported self-recall attempts overall tended to have a higher score on the
delayed test (Appendix A, p. 49).

Let us consider at Memory state 2. Two of the association are dashed to represent being fragile and about to be forgotten. However, by successfully retrieving the visual association between #3 and #4 and #5 and #1, the associations would be strengthened. As long as they are retrievable, they are maintainable.

The forgotten association between #2 and #3 in this memory state could even be repaired, since we can deduce that these objects should be associated since we know that we systematically associated objects to form a loop. One would simply reestablish the association as it was initially created.

On the other hand, repairing Memory state 3 could prove impossible, because object #2 has no intact consciously created associations connecting it to the other objects, and that means that object #2 probably is forgotten due to a lack of associations to it, so the memory state will probably be irreparable.

Now, since the treatment groups’ participants were supposed to form multiple (two) associations to every object, and since the associations might have been relatively rigid (visual), I believe that their memory structures might be quite easy to repair overall. Since their memory structures were as a thread closed to a loop, I imagine it is also very easy to maintain. One would simply follow the thread visually in one’s inner eye. I suggest that perhaps not so many of the participants in the control groups formed memory structures with such ease of maintainability and reparability, and this might contribute explaining the observed results.
Distractions. This is the fourth and final heading on reasons to explain the results. What I now will suggest here could explain why for example Pair Picture and Pair Sketch performed so badly among the treatment groups.

I suggest that the imagery provided in the memorization procedure for Pair Picture, Pair Sketch and Link Sketch actually disturbed the participants rather than helped them overall in the crafting of associations, since Link/Pair Blank without provided imagery got better results. I speculate that the memorization procedure for Pair Picture and Pair Sketch simply had content in too many locations on the screen, in other words that the content was too spread out. The visuospatial sketchpad, our mental visualization capability, is limited, we cannot focus everything at the same time (Goldstein, 2011, pp. 134–136). For example, reading words is easier when the letters are close to each other like in this text. You can consider the problems of having to move your eyes to focus on each letter of the word.

In one of his books, Mattias Ribbing provided scaffolding imagery for practice with the chain method (Method-section, p. 17). In that imagery, he wrote describing text labels tightly to the imagery’s outlines to avoid spreading the content spatially, forcing the eyes to move too much (Ribbing, 2013, p. 15). If I had done something similar, the treatment groups probably would have benefited noticeably, since for them, the object name labels was relatively far from the imagery causing the content to be spread out unnecessarily.

Summary. To summarize, I believe that Link Picture was the only group that benefited by the provided imagery rather than becoming distracted by it. Furthermore, I suggest that Link Picture and Link/Pair Blank’s participants that aced the initial test, which seemed to have formed the most durable memories, did this partially because they successfully managed to create two
rigid visual associations to each object. I suggest that this enabled memory maintenance and repair for a relatively long duration before their memories had degraded beyond repair.

I would now like to suggest how educators could utilize the results of this study, and propose some future research to consider.

5.4 Applications and future research

This study indicate that a written instruction of the Russian doll method efficiently can scaffold memorization of a sequence of objects, which could represent visual retrieval cues to some information of importance, even without providing imagery (Link/Pair Blank), but even better if certain imagery was provided (Link Picture). I believe that users that already have been properly introduced to the Russian doll method will be able to perform well without provided images, because for them the images will not clarify the task further. Moreover, since it requires a notable effort to provide imagery such that Link Picture was provided with, I suggest that such imagery mainly is used while introducing the Russian doll method and that future scaffolding of memorization is done by simply providing written suggestions, as Link/Pair Blank received good results without provided imagery. The benefit of Link Picture was that it caused fewer to cancel the experiment during the memorization compared to Link/Pair Blank, a motivational aspect.

Let me present two examples on how I believe results from this study could be utilized in education, inspired especially by the memory athletes Jonas von Essen, Mattias Ribbing and Idriz Zogaj.

Knowledge outlines. I suggest that teachers and textbook authors provide support or scaffolding to memorize the outlines of the educational content they present. A teacher could then ask students to reflect on the previous lectures outline, who could use the memorized outlines as guides. Motivated students could also study at any time by retrieval practice by
attempting to reconstruct as much of the lecture from the memorized outline, similar to a Greek giving a speech from some memorized notes.

Let me give two arguments for why memorizing knowledge outlines would synergize with retrieval practice. First, retrieval practice depend on successful recalls to be efficient (Kornell et al., 2011), and a knowledge outline might help the person utilizing retrieval practice to successfully retrieve more memories. Secondly, students could study like this during more of their spare time, for example while walking, sitting on the bus or taking a shower, since all they need is some spare concentration. This would lead to more spaced repetition, which have been proven an efficient learning strategy (Goldstein, 2011, Chapter 1).

For a real world application of memorizing knowledge outlines, I have some suggestions. First, I suggest that the educator convey the benefits of retrieval practice and spaced repetition to motivate the students. Secondly, I suggest that students get to know how to use a visual mnemonic method, such as the Russian doll method, for example by testing the memorization procedure of Link Picture (Appendix B).

Now the main point, I suggest that the educator somehow support the memorization of knowledge outlines for the educational content. For example by providing students a table with suggestions on visual retrieval cues for subtopics that cover the content, as I have done for this thesis in Appendix B, Table 7. So if you want, you could memorize my suggested outline for the thesis, and later look back at what you remember about each subtopic, in other words to let it guide some retrieval practice.⁷

---

⁷ Anki ([http://ankisrs.net/](http://ankisrs.net/)) can remind you when to retrieve memories for a spaced repetition, as well as relieve the need to actually memorize the outline, assuming one always have quick access to the mobile.
A perhaps quicker approach to get started with, possible for teachers that are having lectures for students, is suggested by Mattias Ribbing. He suggest teachers to instruct the students to utilize the method of loci, by asking them to visually imagining retrieval cues to a knowledge outline in various locations of the classroom while holding the lecture (2015).

**Catalytic knowledge.** Let me introduce another more general application for the results. Say that you are studying history, and you already know a sequence of related historic events by heart, or say that you are reading international news with an outstanding knowledge of the world map (von Essen, 2014c), or perhaps that you are studying chemistry and know all the relevant details of the elements in the periodic table (von Essen, 2014b). Do you believe that it could help you make sense of some of the educational content, by having such background knowledge available to relate to?

Knowledge can act as a catalyst for learning, and some knowledge, for some educational content, can probably be especially catalytic. Mattias Ribbing, three fold Swedish memory champion, and I separately, suggest that teachers and textbook authors identify the most catalytic knowledge for their subjects and scaffolds students’ memorization of that knowledge (Ribbing, 2013, Chapter 1).

This study does not conclude how to scaffold memorization of any kind of knowledge. It concludes how to help someone remember a sequence a of visual retrieval cues. So anything that can be encoded to a sequence of visual retrieval cues can be memorized this way, but not everything can easily be encoded like that. To scaffold the memorization of for example the periodic table, additional methods are required. For example, to encode the atomic numbers, one
need a system such as the *phonetic number system*, also known as the *major system*, to encode numbers into images.\(^8\)

### 5.5 Conclusions

This study empirically tested if a very quick introduction to the mnemonic method called the Russian doll method (RDM) could support the memorization of twelve objects, which could represent visual retrieval cues to subtopics outlining some educational content. I argue that the results demonstrate that the RDM could be used for this, especially if providing certain imagery for beginners (*Link Picture*), but also without providing any imagery (*Link/Pair Blank*). Participants introduced to the mnemonic method were not only showing a higher overall score, they were also less prone to forget. In addition, some groups’ participants (*Link Picture* and *Link/Pair Blank*) who also initially successfully recalled all objects were clearly less likely to forget than the other participants were. They had formed very durable memories.

I suggested that scaffolding the memorization of an outline of some educational content using the RDM in this way could synergize with retrieval practice, and that scaffolding memorization of catalytic knowledge foundations could prove beneficial.

I believe that the potential of memory athletes’ ideas and methods, in education, not yet have been properly evaluated, nor yet applied to a satisfactory degree. Among the three Swedish memory athletes that I have had some contact with, all believe that increased knowledge about mnemonics would benefit the educational sector (Ribbing, 2011; von Essen, 2014a; Zogaj, 2012).

One of these memory athletes recently received the highest score among those taking the Swedish national university aptitude test, giving him plenty of educational choices, but still

---

\(^8\) A great Swedish introduction to the *major system* is found at: [https://www.youtube.com/watch?v=Hp9KuzoKk9g](https://www.youtube.com/watch?v=Hp9KuzoKk9g)
choose to apply for a teacher education because of the impact he perceives he could have in the educational sector. He is also the current world champion of memory – Jonas von Essen. Another one has earned the title *Grandmaster of memory* and believes that learning the skill to read and write in the memory should be taught in schools just as reading and writing on a paper currently is – Mattias Ribbing (2011). The final memory athlete I have been in contact with, who is a fivefold Swedish champion of memory and manager of the Swedish memory team, believes that:

“[..] this knowledge [about mnemonics], implemented in schools, would change the way we see the school system, not only in Sweden, but in the whole world.”

– Idriz Zogaj (2012)

Please feel free to contact me about this thesis using erik.i.sundell+thesis@gmail.com. It is my goal to create a network especially for teachers, mnemonic athletes and researchers, where we can make collective progress towards evaluating the potential of mnemonics in education.
References


von Essen, J. (2014c). Seterra - all maps successfully completed in one continuous run! Retrieved from [https://www.youtube.com/watch?t=49&v=xBxQTNoGFZQ](https://www.youtube.com/watch?t=49&v=xBxQTNoGFZQ)


Zogaj, I. (2012). *How to become a memory master*. Retrieved from [https://www.youtube.com/watch?v=9ebJlcZMx3c](https://www.youtube.com/watch?v=9ebJlcZMx3c)
HELPING OTHERS REMEMBER

Image references

Appendix A – Additional results

A.1 Experiment durations

Since the experiment was conducted through computer software, it was quite easy to measure the time participant spent on various moments in the experiment. This section summarizes the most measured durations and provides Table 6 that contains the details.

The most important duration measured is the memorization duration. The experiment’s results should be considered with the knowledge of that the control groups’ participants spent less time ($Md = Median = 160 \text{ s}$) on the memorization than the treatment groups ($Md = 240 \text{ s}$). For the control groups, a significant correlation was found between the time spent memorizing and delayed test score, $r(125) = .33$, $p < .001$. For the treatment groups, the correlation was neither strong nor significant, $r(192) = .05$, $p = .494$, probably due to the treatment groups’ low variance of the memorization duration variable.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Web-view name</th>
<th>Sample</th>
<th>n</th>
<th>.05</th>
<th>.25</th>
<th>.50 / Md</th>
<th>.75</th>
<th>.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructions</td>
<td>mem-ins</td>
<td>Control</td>
<td>119</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>15</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td>168</td>
<td>32</td>
<td>55</td>
<td>70</td>
<td>87</td>
<td>150</td>
</tr>
<tr>
<td>Memorization</td>
<td>mem</td>
<td>Control</td>
<td>127</td>
<td>38</td>
<td>97</td>
<td>160</td>
<td>240</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td>194</td>
<td>129</td>
<td>240</td>
<td>240</td>
<td>240</td>
<td>240</td>
</tr>
<tr>
<td>Questions</td>
<td>first-questions</td>
<td>Control</td>
<td>119</td>
<td>35</td>
<td>51</td>
<td>66</td>
<td>87</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td>168</td>
<td>36</td>
<td>55</td>
<td>73</td>
<td>102</td>
<td>207</td>
</tr>
<tr>
<td>Initial test</td>
<td>test</td>
<td>Control</td>
<td>119</td>
<td>51</td>
<td>80</td>
<td>109</td>
<td>171</td>
<td>342</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td>168</td>
<td>50</td>
<td>74</td>
<td>111</td>
<td>157</td>
<td>334</td>
</tr>
<tr>
<td>Test delay</td>
<td>(none)</td>
<td>Control</td>
<td>127</td>
<td>11 h</td>
<td>24 h</td>
<td>24 h</td>
<td>29 h</td>
<td>70 h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td>194</td>
<td>10 h</td>
<td>24 h</td>
<td>25 h</td>
<td>30 h</td>
<td>102 h</td>
</tr>
<tr>
<td>Delayed test</td>
<td>test</td>
<td>Control</td>
<td>117</td>
<td>44</td>
<td>82</td>
<td>112</td>
<td>172</td>
<td>1126</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td>171</td>
<td>47</td>
<td>76</td>
<td>112</td>
<td>155</td>
<td>259</td>
</tr>
</tbody>
</table>

*Note: About 10% of the data in this table was lost due to a technical error.*
A.2 Test score details

**Test score distributions.** The mean test score only gives a course picture of the data. A more fine-grained picture of the data is presented below. Notice especially how the amount of participants acing the test (receiving the maximum score) sets some groups apart.

**Figure 5.** The tests’ score distributions.

- **Figure 5a.** The initial test’s score distributions.
- **Figure 5b.** The delayed test’s score distributions.

**Test acers.** Figure 6 below has isolated the participants that received the maximum score in the tests. We can in it notice that *Link Picture* and *Link/Pair Blank* had formed quite durable memories, as we is presented in detail in Figure 4a-b within the Results-section on page 29.

**Figure 6.** Test acers. The fractions of participants received the best possible score.

- **Figure 6a.** Test acers. Fraction of participants that aced the initial test. *Not enough evidence* was found to reject the differences between the groups as a consequence of random variations; \( \chi^2(7, N = 321) = 8.0, p = .337. \)
- **Figure 6b.** Test acers. Fraction of participants that aced the delayed test. *Enough evidence* was found to reject the differences between the groups as a consequence of random variations; \( \chi^2(7, N = 321) = 19.3, p = .007. \)

*Note: Error bars indicate a 95% CI.*
A.3 Questionnaire data

**Test score estimations.** The participants in the control groups had a stronger confidence in their ability to recall the objects successfully. Participants were asked, directly after the memorization, how much they thought they would score on the initial test and the delayed test. Figure 7 summarizes these estimations.

**Figure 7.** The mean of the participants’ self-estimated test scores. The participants estimated their current and delayed test scores to be significantly higher in the control groups than the treatment groups.

![Figure 7a](Link Pic.)  ![Figure 7b](Link Ske.)

**Figure 7a.** The mean of the participants’ self-estimated test scores of the initial test. This score estimation correlates with a 44% strength with the actual score, \( r(319) = .44, p < .001 \).

The mean of the estimated test scores for the control groups \( (M = 10.6, SD = 1.9) \) was significantly higher than for the treatments groups \( (M = 9.6, SD = 2.4) \).

**Figure 7b.** The mean of the participants’ self-estimated test scores of the delayed test. This estimation correlates with a 35% strength with the actual score, \( r(319) = .35, p < .001 \).

The mean of the estimated test scores for the control groups \( (M = 6.6, SD = 2.6) \) was significantly higher than for the treatments groups \( (M = 7.4, SD = 3.0) \).

*Note:* Error bars indicate a 95% CI.

**Memorization visualization.** In the questionnaire following the memorization, the participants were asked to estimate how much visualizing they incorporated in the memorization on a low-resolution scale with five steps. This measure correlated with the test score of the delayed test, \( r(319) = .35, p < .001 \). The correlation was as strong the participants’ estimation of their score on the delayed test, which means that the degree of visualization was an equally good indicator of the score as the participants’ actual estimation.
Self-recall attempts. After the delayed test, the participants were asked how many different occasions they had tried to recall the memorized objects since they last visited the experiment web page. Having made or not made any self-recall attempts correlated with the delayed test score just a bit weaker than their estimation of the score, $r(319) = .29, p < .001$. 
Appendix B – Experience mnemonic methods

Below are references to two try-it-yourself examples of visual mnemonic methods. For more information, consider contacting me personally (erik.i.sundell+thesis@gmail.com) or joining the Swedish Facebook group “Minnesteknisk i skolan”.

B.1 The Russian doll method

Either use the links below to try the memorization procedure from the group that gained the highest mean test score, *Link Picture*, or read the instructions in Screenshot 5b within Appendix E and then memorize the visual retrieval cues outlining this thesis in Table 7 below.


B.2 The method of loci

If you know Swedish, you can try the memorization of the ten largest countries with instructions given by the current world champion in memory, Jonas von Essen.

- Video link — Memorization: https://www.youtube.com/watch?v=5b5oBdS9dco&t=3m22s (duration: 4 minutes)
- Video link — Test/Interpretation: https://www.youtube.com/watch?v=5b5oBdS9dco&t=13m24s (duration: 3-4 minutes)

B.3 Thesis outline

Memorize the list of visual retrieval cues in the table below as they are encoding an outline of the thesis. Then later, try to recall what the thesis was about, guided by the outline.

**Table 7.** A list of ten visual retrieval cues representing various subtopics covering many parts of the thesis content. Try to memorize the visual retrieval cues using the Russian doll method.

<table>
<thead>
<tr>
<th>Thesis subtopic</th>
<th>Suggested visual retrieval cue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge outlines</td>
<td>Hula hoop</td>
</tr>
<tr>
<td>Mnemonic methods</td>
<td>Russian doll</td>
</tr>
<tr>
<td>Visualization (Large, detailed, colored, in 3D)</td>
<td>Beach ball</td>
</tr>
<tr>
<td>Memory retrieval (retrieval practice, retrieval cues)</td>
<td>Bucket</td>
</tr>
<tr>
<td>Greek speakers</td>
<td>Greek statue</td>
</tr>
<tr>
<td>Willingham’s three memory principles</td>
<td>Stone tablet</td>
</tr>
<tr>
<td>Chunking</td>
<td>Chunk of meat</td>
</tr>
<tr>
<td>Learning styles</td>
<td>Trash can</td>
</tr>
<tr>
<td>Memory maintenance and repair</td>
<td>Wrench</td>
</tr>
</tbody>
</table>
Appendix C – A detailed experiment overview

![Experiment Overview Diagram]

**Figure 8.** An experiment overview. The data collected in part three and part four was disregarded since too few participants completed these parts. The initial task of memorizing twelve objects is the only part that contains group specific treatment variations. The free recall tests measure how well the participant successfully could recollect the twelve objects memorized earlier. To reduce the recency effect (Goldstein, 2011, pp. 151–153), a distraction of questions was used in between the memorization and the first test. Note that the labels within the brackets refer to the different web views on the webpage. See the appendix for a screenshots from each web view.
Appendix D – Participant interactions

Screenshot 1. [information]. The home view of the web experiment.

Link: https://e-sundell.firebaseapp.com/
**Screenshot 2.** Facebook integration. This is how I designed the experiments webpage preview to look within Facebook.

Testa ditt visuella minne
Detta är ett anonynt vetenskapligt webbaserat experiment om visuella minnestekniker. Deltar du får du en personlig rapport om hur du presterade när experimentet är avslutat. Kom och var med! // Erik Sundell - Lärarstudent ©...

E-SUNDELL.FIREBASEAPP.COM
Screenshot 3. The personal result report. The report was sent out after a) the participant completed all parts of the experiment, or b) the participant actively choose to not participate further, which was an option available after completing the first two tests / parts.
Appendix E – Experiment stimuli

All the web-views are provided with links, these links take you straight to the web-view in a debugging mode where no data is saved to the backend. Two of the experiment views available in English, those relating to the memorization procedure.

**Screenshot 4a.** [new]. The first view after going to the experiment area. I state that I wish that the participant a) hasn’t already participated, and b) confirms that he/she owns the given email.

**Screenshot 4b.** [continuing]. The first view after having returned to the experiment as encouraged by an email reminder. The participants are again asked to confirm the ownership of the email.

**Screenshot 5a.** [mem-ins, for the control groups]. Memorization instructions.

**Screenshot 5b.** [mem-ins, for the treatment groups]. Memorization instructions.

**Memorization task**
The goal for this task is to memorize twelve objects so that you later can name the objects, solely by using your memory.

**Choose freely how to memorize the objects**
Do your best! You have four minutes available!

**This will happen**
- You will see two objects at the time.
- Every twenty seconds the objects will change.
- First you will see object #1 and #2, then #3 and #4, then #5 and #6, and so on.

**This is your instruction**
- To remember all twelve object as a shrinking chain.

**Practice example**
1. Imagine a WATERMELON, large as your computer monitor in your mind’s inner eye.
2. Zoom in on an area of the watermelon, and notice additional details in that area.
3. Imagine that a small CUP is attached to the big watermelon.

1. Imagine the CUP again, large as your computer screen in your mind’s inner eye.
2. Zoom in on an area of the cup, and notice additional details in that area.
3. Imagine that a small MICROWAVE OVEN is attached to the big cup.

Link (Swedish): [https://e-sundell.firebaseapp.com/#/experiment ...](https://e-sundell.firebaseapp.com/#/experiment ...)

Link (English): [https://e-sundell.firebaseapp.com/#/experiment ...](https://e-sundell.firebaseapp.com/#/experiment ...)
**Screenshot 6 [a-i].** [mem]. The memorization procedure for the eight different groups. An English translation is available.

<table>
<thead>
<tr>
<th>Picture</th>
<th>Sketch</th>
<th>Blank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Picture <em>(Eng, Swe)</em></td>
<td>Control Sketch <em>(Eng, Swe)</em></td>
<td>Control Blank <em>(Eng, Swe)</em></td>
</tr>
<tr>
<td>Control Sketch <em>(Eng, Swe)</em></td>
<td>Link/Pair Blank <em>(Eng, Swe)</em></td>
<td></td>
</tr>
</tbody>
</table>

- **Link Picture *(Eng, Swe)*:**
  - Pingisracket
  - Guitar

- **Pair Picture *(Eng, Swe)*:**
  - Pingisracket
  - Guitar
Screenshot 7a. [first-questions]. These questions were asked directly after the memorization procedure, with the aim to reduce the *recency effect*, which causes recently occurred events to be easier to recollect. The questions were about: a) age, b) sex, c) geographical location, d and e) estimation of the test score received right now and if tested in 24 hours, f) previous experience with visual mnemonics, g) about how living the visualization became in the mind’s inner eye during the memorization procedure, and h) a free comment about the memorization procedure.

**Några korta frågor**

<table>
<thead>
<tr>
<th>Ålder (år)</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kön</td>
<td>Könna</td>
</tr>
<tr>
<td>Bostadskommun</td>
<td>Uppsala</td>
</tr>
<tr>
<td>Poänguppskattning nu</td>
<td>12</td>
</tr>
<tr>
<td>Poänguppskattning om 24 timmar</td>
<td>12</td>
</tr>
<tr>
<td>Visuell minnesteknisk erfarenhet</td>
<td>Ja</td>
</tr>
<tr>
<td>Visuell inlevelse</td>
<td>5 - Mycket stor</td>
</tr>
<tr>
<td>Fri kommentar om memoreringen</td>
<td>Vill du skriva en kommentar om memoreringen kan du göra det här.</td>
</tr>
</tbody>
</table>

Link: [https://e-sundell.firebaseapp.com/#/experiment/?debug=true&stage=1&group=2&route=first-questions](https://e-sundell.firebaseapp.com/#/experiment/?debug=true&stage=1&group=2&route=first-questions)
Screenshot 7b. [mid-questions]. These questions were asked to the participants that had just finished a delayed test. The first question is about self-recall attempts, and the second about sleep. Self-recall attempts: “At approximately how many different occasions have you attempted to recall the images, during the time that elapsed between finishing the last experiment part, until you started this part?”. Sleep: “I have slept since the latest test” or “I have not slept since the latest test”.

Screenshot 7c. [last-questions]. These questions were asked to the participants that were about to finish the fourth part of the experiment. They stated if they wanted one email after the work was completed with exiting material about visual mnemonic methods, as well as give free comments.
Screenshot 8a. [test-ins]. This view contained brief instructions for the free recall test to come. The instructions were: a) Write down every object you can recall, b) Don’t worry about the spelling, c) The order does not matter, and d) Only use your memory.

Link: https://e-sundell.firebaseapp.com/#/experiment/test-ins?debug=true&stage=1&group=2&route=test-ins

Screenshot 8b. [test, for the control groups]. The input fields were positioned in a 3 x 4 grid to match how they memorized the objects.

Link: https://e-sundell.firebaseapp.com/#/experiment/...

Screenshot 8c. [test, for the treatment groups].

Link: https://e-sundell.firebaseapp.com/#/experiment/...
Screenshot 9a. [return-info, part one]. This view was shown after finishing the experiments first part.

Screenshot 9b. [return-info, part two]. This view was shown after having returned at least once to make a delayed free recall test and done that.

Screenshot 9c. [done]. The view shown after completing the full experiment. The YouTube video shown in this view is a presentation by the Swedish world champion in memory, Jonas von Essen. The address to that video is https://www.youtube.com/watch?v=5b5oBdS9dco

Link: https://e-sundell.firebaseapp.com/#/experiment/?debug=true&stage=8&group=2&route=done
Appendix F – Experiment administration

Screenshot 10. The test answer categorization tool. All unique answers were manually categorized as either a) wrong or b) correctly representing a specific object using this custom tool.

Screenshot 11. The control view of the experiment administration spreadsheet. It contains a summary of the experiment status and settings associated with the experiment administration. A custom menu bar with quick access to custom commands was also created, as seen in this screenshot with the label “Experiment Administration”.

Screenshot 12a. The administration log. It logs experiment activity and exceptions thrown in the administration code. Administration scripts were automated to run every thirty minutes. One of the scripts' tasks was to send out email reminders asking the participant to return for another test.

Screenshot 12b. The client-side log. It logs errors from the participant's web browsers. This view turned out to be very useful and important. Some bugs were identified and corrected using this log. The entries were logged by a Google Apps Script, which listened to error reports sent from participants' web browsers.

Screenshot 13. An initial analysis. Initially I used my knowledge of Google Spreadsheets to do a preliminary analysis of the collected data as it was collected. Later I decided to use the statistical program STATA instead, programming nearly one thousand lines of code to plot graphs and calculate statistics.