



**University of  
Zurich<sup>UZH</sup>**

Master's thesis  
presented to the Faculty of Arts  
of the University of Zurich  
for the degree of  
**Master of Science in Psychology UZH**

# **Analogical Transfer in a Hebb Repetition Paradigm**

**Author: Oesch Adrian**  
Student ID Nr.: (09-724-063)

Examiner: Prof. Klaus Oberauer

Department of Psychology

Submission date: 20.09.2016

### **Abstract**

Previous research has shown unintentional analogical mapping processes due to relational similarity when participants process word pairs (Estes & Jones, 2006; Popov & Hristova, 2015). Also, it has been shown that structural similarity guides analogical retrieval in reminding studies (Wharton, Holyoak, & Lange, 1996). This thesis aims to combine those lines of research and investigates unintentional analogical transfer on the basis of mere structural similarity. By combining an established method from memory research with original story material, we developed a novel experimental setting to investigate this hypothesis and contribute to growing evidence that analogical transfer is not a deliberative reasoning process but happens also unintentionally. We detected weak evidence (5%-sign. level) for this effect in experiment 1, where we observed an increased recall performance over trial index in stories with high structural similarity compared to stories with low structural similarity. This effect could not be replicated in experiment 2, where we changed the trial sequence to a repeated condition setting and introduced limited reading time. We then discuss the differences of those two experiments, further methodological improvements and possible future experimental endeavors.

## Contents

<b>Introduction</b>	<b>5</b>
Analogy: What is it good for? . . . . .	5
Analogical Transfer and Problem Solving . . . . .	7
Unintentional Analogical Processes . . . . .	8
Structural Similarity and Analogical Access . . . . .	10
Hebb Repetition Paradigm . . . . .	11
 <b>Experiment 1</b>	 <b>13</b>
Methods . . . . .	13
Task . . . . .	13
Participants . . . . .	14
Material . . . . .	15
Procedure . . . . .	17
Results and Discussion . . . . .	18
Material Assessment . . . . .	18
General Task Performance . . . . .	18
Confounding Effects . . . . .	19
Main Effect . . . . .	22
Reading Time . . . . .	25
Follow-Up . . . . .	27
Conclusion . . . . .	28
 <b>Experiment 2</b>	 <b>29</b>
Methods . . . . .	29
Task . . . . .	29
Participants . . . . .	29
Material . . . . .	29
Procedure . . . . .	30
Results and Discussion . . . . .	30

ANALOGICAL TRANSFER IN A HEBB REPETITION PARADIGM	4
Material Assessment . . . . .	30
General Task Performance . . . . .	30
Confounding Effects . . . . .	31
Main Effects . . . . .	33
Follow-Up . . . . .	36
Conclusion . . . . .	36
<b>General Discussion and Future Directions</b>	<b>38</b>
Material . . . . .	38
Processing Depth . . . . .	38
Task difficulty . . . . .	40
<b>Summary</b>	<b>41</b>
<b>References</b>	<b>42</b>
<b>Appendix</b>	<b>46</b>
Material Experiment 1 . . . . .	46
Instructions on CrowdFlower . . . . .	46
Follow-Up Questions . . . . .	46
Stories with Relation Response Options . . . . .	46
Names . . . . .	48
Material Experiment 2 . . . . .	48
Instructions on CrowdFlower . . . . .	48
Stories with Relation Response Options . . . . .	49

## Introduction

### **Analogy: What is it good for?**

Recently while listening to a podcast about economics, an expert compared money to blood. He explained that when our economy would be our body then the blood would be the money and the heart would be the financial sector distributing the blood around the body. I immediately felt I had a better understanding of why money and banks are relevant to our financial system. This is a typical example of analogy. Through the alignment of entities in a system we don't understand to the entities in a better known system, we start attaching the unknown to the known. Every analogy has two models, a source and a target model. Because the expert tried to retrieve information from the better known model, i.e. the blood system in the previous example, the blood model is the source and the monetary system is the target. People can draw inferences from one model to the other, that's why it is also called analogical transfer and the process of alignment is called mapping. In a "perfect" analogy every element of the source model has a counterpart in the target model. Blood gets mapped on to money and the heart on to financial system for example. The better the entities of two models match on to each other, the higher systematicity of the analogy is (Gentner & Holyoak, 1997). We can try to align more elements to increase the systematicity of our previous example: If blood is money, and the heart is the financial sector, the organs might be different industries, each industry has a different job to do, but still all of them rely on having a big enough blood supply. Also, the higher the blood supply in an organ is, the more work it can get done. This could be interpreted with higher investment opportunities, in some organs than in others. The brain for example uses a lot more oxygen and therefore needs a higher blood supply, but then again it is probably worth it. Analogically one could say, that the insurance industry is such a finance heavy sector that is very important to an economy. But usually this alignment process can also be stretched too far, and the mapping opportunities have limits. If blood is money, what then is the oxygen in the blood regarding the financial sector? Generally people tend to prefer analogies with higher systematicity for inference as the coherence and soundness

of the comparison is increased (Clement & Gentner, 1991).

Due to its powerful applications analogies can be found in numerous professional domains. An often cited example within the realm of schools and education is the comparison of electricity to the flow of fluids that teachers use to explain new concepts like voltage in science class (Dupin & Joshua, 1989; Gentner, 1982; Treagust, Duit, Joslin, & Lindauer, 1992). Other researchers have investigated analogies in politics. It made a lot of sense for president Bush Sr. to compare Saddam Hussein to Adolf Hitler when entering the war in Kuwait in 1991, because he wanted the attributes from Hitler to be also attributed to Saddam Hussein (MacDonald, 2002). Analogies are also common in other situations of persuasion like in marketing (Cornelissen, 2003; Herzenstein & Hoeffler, 2014). Analogies are also considered to be very important in the innovative process in science, research and development (Gentner, 2002). Dunbar (2001) found that scientist heavily rely on analogies in lab meetings. Yet another impressive feature of analogies is that knowledge that was learned through analogical comparison seems to be used more frequently in similar subsequent tasks, then if the same material is just summarised (Gentner, Loewenstein, & Thompson, 2003; Loewenstein, Thompson, & Gentner, 1999). And it is also not a coincidence that the Ravens Matrices Test, which heavily relies on analogical reasoning, is one of the common standard tests to measure general intelligence (Carpenter, Just, & Shell, 1990).

Apart from systematicity, analogies can also differ regarding their the level of similarity. Comparing a story with dogs to a story with cats is a close analogy, because both animals share a lot of surface features like the amount of legs, tails, fur, way of moving and so on. Contrary to comparing a story with a dog to a story with people, one has to focus on higher order relations to detect similarities, which allow mappings from one model to the other. Without any kind of similarity, it is not possible to map models on to each other, because then no entities can be aligned. Remote analogies - the mapping of models with no or only little surface similarity - are the champions league of analogies, others call it true analogies (Wharton et al., 1996), because they rely merely on structural similarity or higher order relational similarity (Catrambone, 2002) and

therefore are harder to detect, as we will show later in this thesis. Detecting remote analogies can be especially advantageous in problem solving situations, when search strategies based on surface features have not triggered a detection of a suitable solution. One can also think of it as extending the search space. By relying on abstract features, the amount of potential candidates for comparisons one could make increases a lot.

### **Analogical Transfer and Problem Solving**

In a famous experiment M. Gick and Holyak (1980) first presented participants three stories. Participants had to first memorize those stories and recall the content after each story as precisely as possible. One of those stories was about an army general, who wanted to invade a fortress. The general knew that if his army reaches the fortress, his force will be big enough capture it, but there was a problem. The army couldn't simply march onto the fortress, because the road leading to the fortress was mined and only small groups could safely pass it. The solution to the general's problem was to split his army into multiple smaller forces and let them simultaneously march on multiple roads from different directions onto the fortress. That way the many smaller groups could pass the mined roads and rejoin forces at the fortress, to then attack and conquer it. it is also called an attack-dispersion solution. The other two distractor stories were as disanalog as possible to the general story but similar in length.

After all stories were memorized, recalled and a short break, participants were presented with a new a problem-solving task: A doctor has a patient with a tumor in his stomach. The patient can't be operated but is going to die unless the tumor is removed. The doctor could use kind of a ray. The rays could destroy the tumor when intense enough, but they would also damage the healthy tissue surrounding the tumor. At lower intensity, the ray doesn't damage the healthy tissue but neither the tumor. The solution to the doctor's problem is, to apply the attack-dispersion schema like in the story with the army general: Split the bundled force into multiple weaker ones, and only join the forces together where it is necessary.

The way the two stories are presented in this thesis, mapping the general's story as the

source to the doctor's problem as the target story might be obvious. But in the experiment by M. Gick and Holyak (1980) only 21% of participants in the group without a hint came up with the dispersion solution, versus 91 % in the group with a hint. M. L. Gick and Holyoak (1983) replicated the solution rates in various slight modifications of this experiment, leaving the question open: Why is it that hard to detect similarities between systems based on structural similarity? Why do spontaneous remote analogies occur relatively rarely?

M. Gick and Holyak (1980) argue that memory might be sensitive to the encoding context. In the first experimental phase the story material was encoded in the the context of memorization task, while later they were faced with a problem solution task. Why should anyone assume those tasks have anything in common and therefore relate the content of both task to each other? Also, they further argue that analogical retrieval might be sensitive to the semantic context of the target story in problem solving tasks. Why should one be reminded of an army general when trying to solve a doctors problem? One could just as well think of farmers or teachers. They share about the same features (surface similarity) with the doctor as the army general. And indeed, various experiments showed that analogical retrieval is much more likely when the source and target models have similarities on lower levels of similarity e.g. the surface (Gentner, Rattermann, & Forbus, 1993; Holyoak & Koh, 1987).

### **Unintentional Analogical Processes**

In contrast to the line of research provided in problem solving paradigms, more recent studies have started to investigate analogical processes that happen on an much more frequent basis. While in the standard problem solving paradigm spontaneous analogical transfer only rarely occurs, studies involving relational priming investigate whether analogical processes might be effective in much more general cognitive processes. Priming studies have been widely used to study semantic memory throughout the second half of the 20th century (Lucas, 2000), but rarely have they focused on the relations between objects. Processing of relations between objects is crucial to the



mapping process in remote analogies, because the detection of similarity between two remote systems can only rely on structural similarity. Relations between objects can be viewed as a first level of abstraction Catrambone (2002). Spellman, Holyoak, and Morrison (2001) conducted one of the early relational priming studies. They found that participants were faster in lexical decision tasks when the same relation was implied, but not explicitly mentioned, by a previous pair of words, but only when participants were instructed to note and use the relations. The effect was not observed when subjects were merely reading the word pairs. It has been argued, that those effects were mere side-effects from associative priming (Gagné, Spalding, & Ji, 2005), but Estes and Jones (2006) presented evidence, that this is not the case, by controlling for lexical similarity independently from relational similarity.

Hristova (2009) additionally could show that relational priming is also present in a color naming paradigm with pairs of words, where one word per pair changed its color 1'000 milliseconds after appearance. When the prime word pair was in green and the cue word pair also turned green, therefore the colors were congruent to each other, then the participants reaction times were faster when those two word pairs were analog, suggesting that an unintentional mapping process facilitates the color naming process in congruent color situations, resulting in shorter reaction times and even though people were not asked to relate the content of the word pairs to each other.

Day and Gentner (2007) applied a different experimental setting to uncover unintentional analog processes. In this experiment participants were asked to read two stories, after the second story they were asked a few open ended question, which were not clearly answerable from the previous story itself. Two groups differed in terms of the first story that they had to read, which differed in terms of levels of similarity. As dependent variable the investigators checked the inference that the different groups made from the first story to the interpretation of the second story. The two groups differed significantly in terms of how they interpreted the second story, depending on which story they had read first. They were not instructed to do so, and when asked in a post-experiment survey, almost 80-% of participants were responding, that they felt

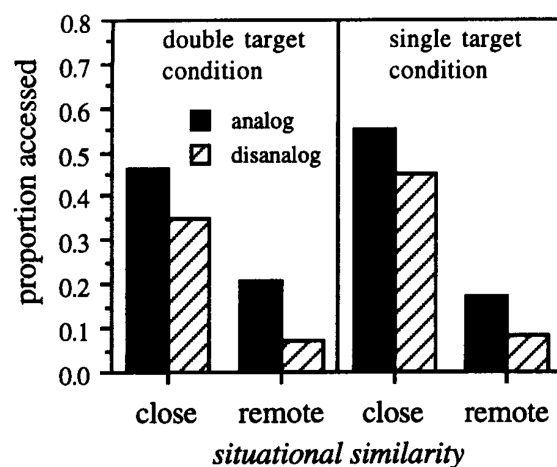
that the stories were understandable by its own and without inference from other stories. In addition to this finding, the participants also encoded the target statements faster, if the statements were systematically more consistent to the source.

These experiments go along with a stream of thought, which views analogy making as a much more fundamental cognitive process than just in executive functions like reasoning. Chalmers, French, and Hofstadter (1992) describe analogical reasoning as kind of a high level perception. This theory proposes a universal process of information aggregation throughout different levels of perception. Remote analogies, the main topic of this thesis might be only the tip of the iceberg. Also, D. R. Hofstadter (2001) argues that analogical reasoning is the actual core of cognition, "for without concepts there can be no thought, and without analogies there can be no concepts" (D. Hofstadter & Sander, 2013, p. 34). The role of analogies as a fundamental cognitive process has already been expressed by other researchers as well. While not under those terms Kokinov (1994) developed a model he called "Associative Memory-Based Reasoning" where they tried to connect analogical reasoning literature to other streams of cognition research. Others argue, that relational priming is the basis of analogical access, relying mainly on research from developmental psychology (Leech, Mareschal, & Cooper, 2008). Now so far various experiments show the rare spontaneous analogical transfer based on structural similarity in problem solving, while at the same time it has been shown, that models are also being unintentionally mapped on to each other on the foundation of mere relational similarity. Taking the current state of research regarding unintentional analogical transfer one step further, this thesis aims to find evidence for unintentional analogical transfer on the foundation of mere structural similarity.

### **Structural Similarity and Analogical Access**

Holyoak and Koh (1987), Johnson and Seifert (1992) and Wharton et al. (1996) presented evidence that structural similarity has an effect on analogical reminding. In the experiments by Wharton et al. (1996) participants were asked to read a set of target stories first. After a certain time, which they varied across experiments, the participants

were asked to read another set of cue stories. Each of those cue stories had either one or two potential target candidates, depending on the condition double target or single target. They varied the relations between cue and target stories according to two factors: surface similarity and structural similarity (they used situational similarity and thematic similarity for those terms). In the double target condition both target stories were either remote or close, but differed in their structural similarity. Figure 1 depicts the result of this study. At first, one can detect a clear difference between close and remote cue stories, confirming previous experiments on the significance of shared low order features. But when comparing the analog to the disanalog cue stories, those experiments also indicate an effect of structural similarity on analogical reminding, therefore the authors conclude that structural information does guide analogical reminding. The authors although emphasize that reminding doesn't necessarily lead to analogical retrieval, as the instructions in a reminding paradigm differ crucially.



*Figure 1.* Reminding frequencies from Wharton et al. (1996) for both close/remote (surface similarity) and analog/disanalog stories (structural similarity).

### Hebb Repetition Paradigm

To test the hypothesis whether structural similarity suffices for unintentional analogical transfer, we applied a experimental paradigm from memory research. The Hebb

repetition paradigm was first introduced by Donald Hebb in 1963 (Lafond, Tremblay, & Parmentier, 2010). He presented participants a few different wordlists while participants were asked to recall the items on those lists after each list. One specific wordlist was repeated every third trial. The Hebb repetition effect describes the increase in recall performance over repeated presentations of the same wordlist. This effect has been studied since its introduction in numerous ways with variations regarding the size of the wordlist, the content of the wordlist, the amount of filler lists between the target (repeating) list and complexity of the task (Lafond et al., 2010; Oberauer, Jones, & Lewandowsky, 2015). Hebb and other researchers saw this experimental paradigm as a model of long term learning because the effects of increased memory performance on the target list often lasted a long time. Due to the thorough investigations of this effect and robust results, we think the Hebb repetition paradigm is a sensitive enough candidate, to test our hypotheses. Instead of presenting wordlists we constructed two pools of stories. A pool with analog stories and one with disanalog stories. Through the repetition of the causal structure of a story, we want to replicate the increased memory performance effect as seen in the Hebb repetition paradigm. Especially the possibility of repeated activation of an abstract structure in this setting, seem to make it a suitable candidate.

## Experiment 1

### Methods

All experiment material source code is available on Github (Oesch, 2016).

**Task.** To test those hypotheses we set up an online experiment using free hosting resources on Openshift (<http://www.openshift.com>). To run the experiment in the browser the open source JavaScript library jsPsych (de Leeuw, 2015) was used, which was developed explicitly to conduct psychology experiments in the browser. Similar to the Hebb Repetition Paradigm, we developed a novel task, where participants had to read a sequence of very short stories and recall the content after each story. All stories differed on the level of the surface, meaning that the semantic context of the story differed between all stories. Every second story had an analog story structure, resulting in a sequence of alternating analog and filler stories. Each story consisted of three statements. We piloted the length of stories, so that the task wasn't too difficult on the one hand, but also not too hard to come up with suitable stories on the other hand. After the presentation of a fixation cross for 500ms, a story was presented statement by statement. Participants were free to choose when to continue to the next statement by a forward arrow key press, but they could not go backwards. We decided to opt for free reading time, to be able to check if a repeated analog story structure might also have an effect on encoding, as was observed in Day and Gentner (2007), or the reading time might also have an effect on recall success. We mention in the instructions to "read the statements thoroughly, but don't waste too much time". While the first two statements each had the structure: entity, relation, entity. The last statements only consisted of an entity followed by a relation, which represented a conclusion from the two previous statements. After each story, the participants were asked to recall the content of the story, again statement by statement by choosing the correct name of the entities and relation out of two separate menus (one menu for the names and one for the relations). Figure 2 depicts an example of the response display. Participants could use drag and drop to reconstruct the statements with the correct items. They were allowed to use as many drag and drop moves as they want and did not have a time limit either when

reading or reconstructing a statement. Figure 3 illustrates the sequence of displays of a single story/trial.

Menu:

Names:	Relations:
Upel	has a lot of
Pishi	is sparked through
Kesal	is rarely found in
Eyem	fills in for
Ahow	starts trading
Vilux	starts brokering
	is filled with
	starts negotiating
	is sparse in

Response: **Statement 1**

Name:	Relation:	Name:

Next

Figure 2. Example of a response display for the statement "Mining company Ahow has a lot of resource Pishi."

**Participants.** For data collection we collected 109 observations from the online labour platform CrowdFlower. In the description of the task on CrowdFlower we explicitly asked for native english speaker, and set the country parameter on CrowdFlower to United States only. Nonetheless a few participants responded with different native languages than english in our post experiment survey and were therefore excluded from further analysis. In addition, 17 participants didn't pass the trick story, where they were supposed to recall a single statement (2 names and one relation) and

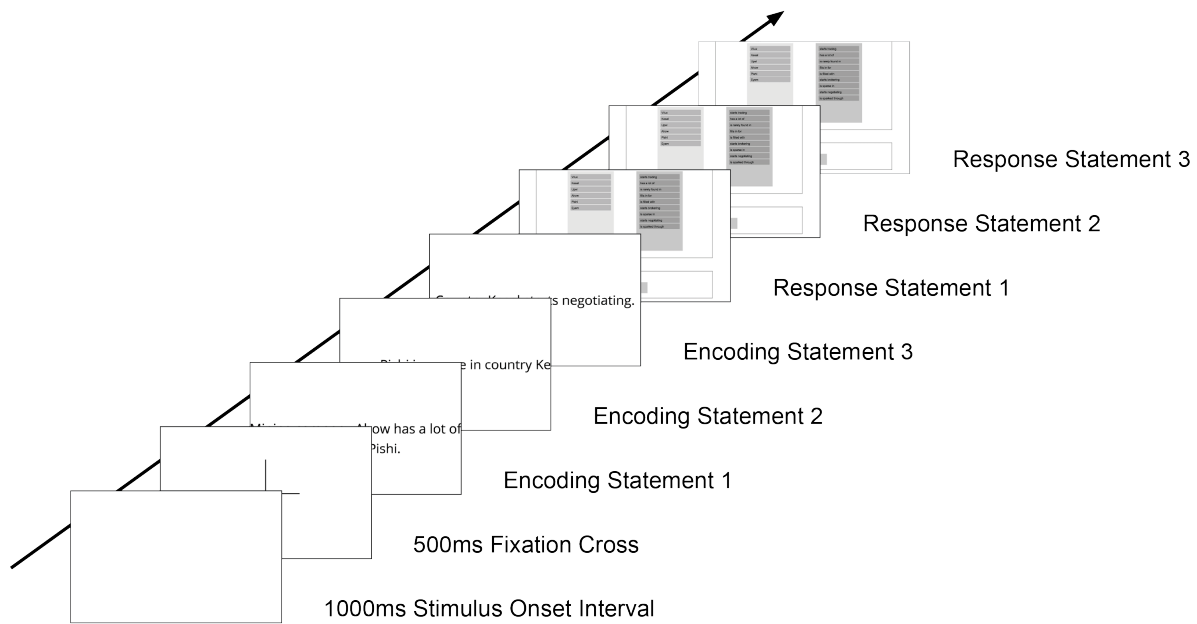


Figure 3. Sequence of a single trial.

where the response options offered 2 names and 3 relations. Those subjects were also excluded, which resulted in a sample of 85 subjects. Figure 4 shows the sample's age distribution which varied between between 20 and 65 years old (median at 37), a rather wide age spectrum compared to traditional college student samples. The sample was equally distributed in terms of gender (50.3% female). According to the self-reported highest degree of education, most of our sample had a university/college degree (58%). Fewer had a high-school degree (39%), and three subjects (3.6%) reported having having a doctoral degree.

**Material.** We constructed two sets of short analog and filler stories consisting of three statements each. The goal was to build all stories with a minimum of surface similarity overlap, i.e. no two stories with animals or where plumbers have a lead role. Three named entities were part of each story. To rule out individual differences in name recognition, and also to make the task more difficult, the names were randomly constructed by a free online service on dialectcreator.com so they were nonsensical. All analog stories had the following causal structure: "Entity A is responsible for entity B",

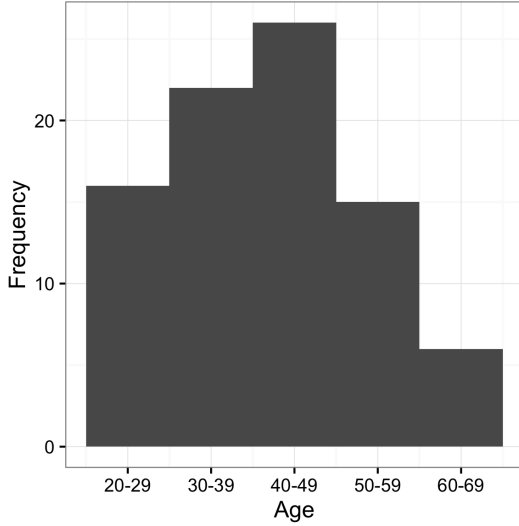


Figure 4. Distribution of age in the sample of experiment 1.

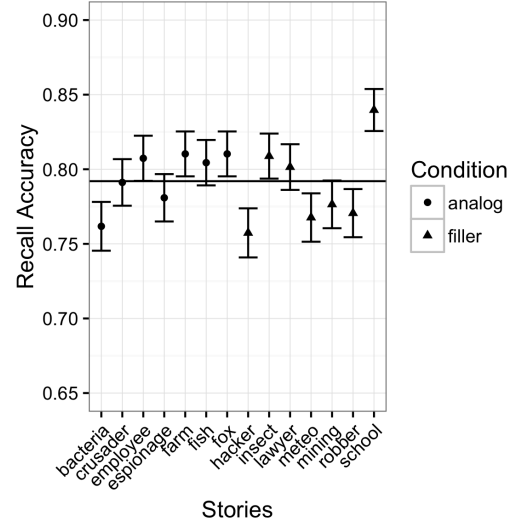


Figure 5. Mean recall accuracy and standard error for each story plus grand mean.

"Entity B overcomes protection of entity C" and "Entity C endures negative consequence". The filler stories were constructed to be linguistically as similar as possible and therefore the sequence of relations and entities resembled the analog stories. The major difference between analog and fillers stories was the lack of systematicity when trying to map them on to each other, due to the lack of a shared higher order relational structure. To see the full table of stories and names see the appendix. As response options we offered another alternative for each name, resulting in a total of 6 name-options per story, while only 3 of those appeared in the presented statements. For each relation from a story, we constructed two alternative response options, one was supposed to be synonymous to the correct relation, and one which was linguistically congruent but not in terms of the meaning. This resulted in a total of 9 relation options per story. For each story we were recording 3 relation responses and 5 name responses (2+2+1 per statement), resulting in 9 items per story. We constructed 7 analog and 7 filler stories for this experiment. This gave us  $85 \times 14 \times 8 = 9520$  binary observations of recall success to work with.



**Procedure.** Participants had to open a link on CrowdFlower where they were sent to a webpage on our server on Openshift. We asked them to shut all other programs so they weren't disturbed by notifications. They were also reminded to put themselves in a quiet environment and reduce distractions like noise to a minimum. The experiment then launched into fullscreen mode and after giving consent and reading the instructions, the memory task was started. Alternately seven analog and filler stories were presented. The order within the analog and filler stories was randomized. Participants either started with an analog or a filler story, although we had to fix this part of randomization due to a bug after we started collecting data. Subjects starting with filler stories were therefore collected at another date. After the 12th story we injected a trick story consisting of only one statement and two names, to check if the participants were paying attention. As response options, only the two names which appeared in the previous statement and three relation options were presented. After the trick story, we resumed with alternating another set of analog and filler stories.

Following the memory task, a set of three closed yes or no questions regarding the conscious registration of different aspects of repetition were asked for, with the options to explain in an open format if participants responded yes to the closed question. The questions built up on each other, each being a little more specific about the repeated causal structure than the previous one. The first one asked if participants noticed any similarities between the stories, while the second one asked for repeating elements between the stories and the third question explicitly mentioned the repeated analog/causal structure. See the appendix for exact wording of the questions.

Thereafter a set of questions regarding the demographics (age, gender, highest degree of education, native language) followed. Additionally we asked for the amount of effort on a four point Likert scale ("0 - none" / "1 - a little bit" / "2 - quite some" / "3 - a lot"). We also asked if they completed the experiment in a serious manner and we therefore should use their data for analysis. We promised to give them the regular payment independently of their answer to this question. For the exact wording of the instructions used on CrowdFlower and in the experiment please see the appendix or the

raw material on Github (Oesch, 2016). At the end of the experiment, every participant got a confirmation code, which they had to paste back into the form on CrowdFlower. After completion of the task Participants were paid one dollar for going through the whole experiment and received a bonus of another two dollars if they showed effort. The median of total experiment duration was 21 mins. Due to continuous bug monitoring, data was collected over various dates.

## Results and Discussion

The raw code of the scripts used for figures and analyses is online on Github (Oesch, 2016).

**Material Assessment.** Even though we tried each story to be equally difficult to memorize, when dealing with language material it is hard to neutralize a bias introduced by language. Figure 5 indicates a certain difference regarding story difficulty, which is confirmed in an analysis of variance on item level,  $F(13,9506) = 2.303$ ,  $p=0.005$ . It is therefore important that we account for this bias in further analyses with random intercepts.

**General Task Performance.** General recall accuracy was much higher than expected. A pilot study investigating suitable story length resulted in overall recall accuracies around 60%. Median of overall recall accuracy in this experiment was at 83% and over 30% of our participants had recalled more than 90% of items correctly. Figure 21 depicts the distribution of overall recall accuracies per subject. We mentioned in the instructions, that no means to facilitate this task were allowed, contrary to the pilot study, where we didn't emphasize this. Also, this time we incentivized subjects to put effort into the task so they will receive a bonus, which could have led subjects to cheat. When comparing the median statement reading time per subject with the overall recall accuracies per subject (see Figure 7), one can see an improved recall accuracy with increased median encoding time, the correlation is significant ( $r(83) = 0.55$ ,  $p < 0.001$ ). Half of our sample has a median of more than 6.5 seconds reading time for a single statement. This could be enough time to take notes (a little over 2 seconds per item),

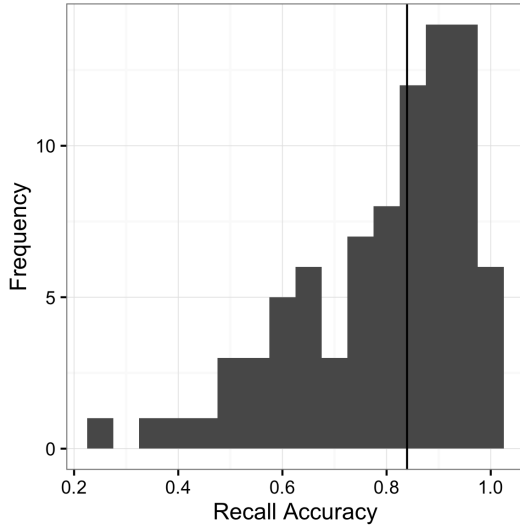


Figure 6. Distribution of overall recall accuracies per subject with median.

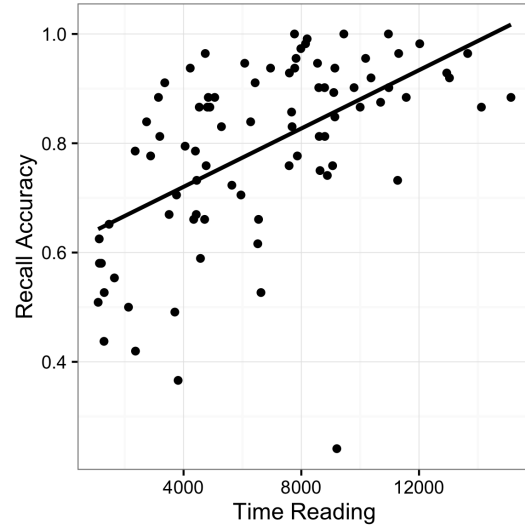


Figure 7. Correlation of subject's median statement reading time in milliseconds with overall recall accuracy.

or just really put a lot of effort into encoding those two or three items and repeating the previous statement in a loop. Although when comparing self reported effort with recall accuracies then this does not seem to have a big effect either on recall accuracy (see figure 8) and encoding time (see Figure 9). An analysis of variance with recall accuracy as dependent variable and effort as independent variable was not significant ( $F(1,83) = 0.24, p = 0.626$ ). The ANOVA with the subject's median reading time was neither significant ( $F(1,83) = 0.83, p = 0.364$ ).

All in all those observations lead to the conclusion, that the task didn't reproduce a very natural reading and text comprehension environment. Participants did take a lot of time into encode or rehearse the items and we can not exclude the fact, that participants cheated by taking notes for example.

**Confounding Effects.** Nonetheless we could reproduce a set of well known effects from cognition literature with this setup, which legitimizes the interpretation of our results to a certain extent.

At first, we observed a general training effect as shown in figure 10. A univariate logistic regression predicting recall success with the trial index is highly significant ( $e = 0.043$ ,

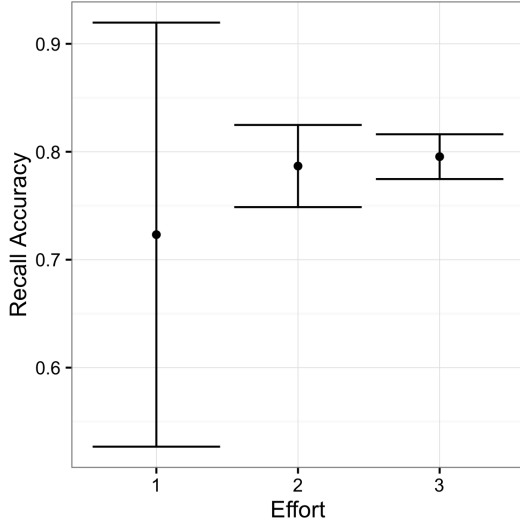


Figure 8. Self reported task effort on a 4-item Lickert scale (zeros removed) with mean and standard errors of recall accuracies per subject.

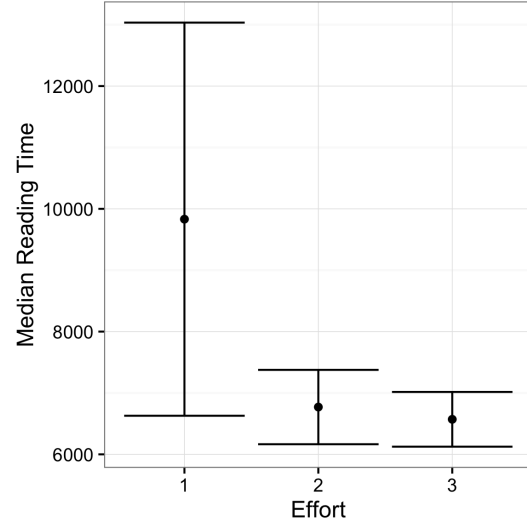


Figure 9. Self reported task effort (see Fig. 8) with mean and standard errors of per subject median reading time.

$z(9518) = 7.321, p < 0.001$ ). Although the trials 1 and 12 seem to deviate quite strongly from the general trend. It could be the case, that even though we presented participants with a pilot statement to give them the option the get accustomed to the response menu beforehand, participants didn't fully understand the task until after the first full trial. Further it should be noted that the trick story was inserted at the 11th trial position, but of course excluded from this analysis. Apparently the lack of coherence regarding story length confused participants remarkably and is observable in recall success.

Each statement had both entity names and relations. Figure ?? depicts the difference in recall performance for each type of response item. The two different name responses did not differ significantly ( $\chi^2(1, N = 5950) = 1.219, p = 0.270$ ), which is why we grouped both name responses together for further analyses. Relation items were significantly easier to recall ( $\chi^2(1, N = 9520) = 113.78, p < 0.001$ ). This effect can most probably attributed to the nonsensicality of the names, which makes them harder to recall, but it can also be argued that participants had to recall 5 names while only 3 relations. Maybe storage resources weren't equally distributed, but we will not go further into this topic.

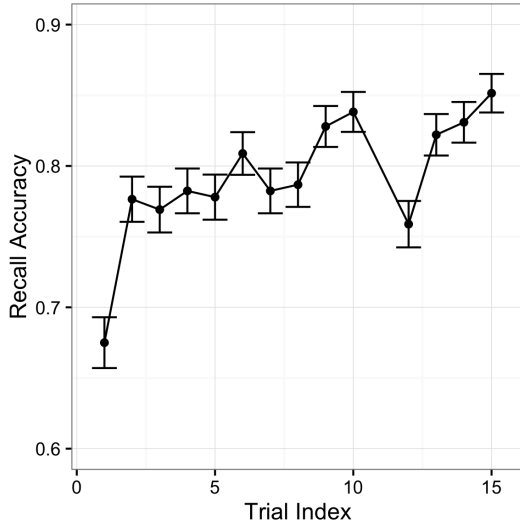


Figure 10. Mean recall accuracy and standard errors for each trial index.

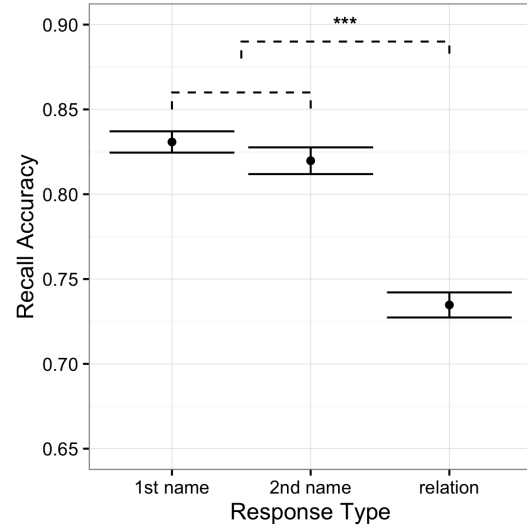


Figure 11. Mean recall accuracy and standard errors for each response type.

We already analyzed the influence of median reading times per subject and detected its effect on a subjects overall recall accuracy (see General Task Performance). Therefore it is not surprising to see, that also in an item level the simple logistic regression analysis, the regression coefficient from the logarithm of the statement's reading time is highly significant ( $e = 0.519$ ,  $z(9518) = 18.88$ ,  $p < 0.001$ ).

As each story had three statements, we were interested, if we can replicate a primacy and recency effect. Figure ?? shows a clear difference in recall accuracy depending on the statements position in the story. A logistic regression predicting recall success with two dummy variables confirmed the statistical significance for the comparison second statement vs. first statement ( $e = -0.233$ ,  $z(9517) = -4.06$ ,  $p < 0.001$ ), and the comparison 3rd statement vs. 1st statement ( $e = 0.153$ ,  $(z(9617) = 2.24$ ,  $p = 0.025$ ). Although it should be noted, the second statement was linguistically the most complex, and the last statement in each story only had two items to remember. So it is not clear if the significant differences can be attributed to primacy/recency or rather to differing statement difficulties.

The two pools of stories we constructed, analog and fillers, only differed very slightly in terms of recall difficulty (see figure 13). The mean of recall success only differs 2%

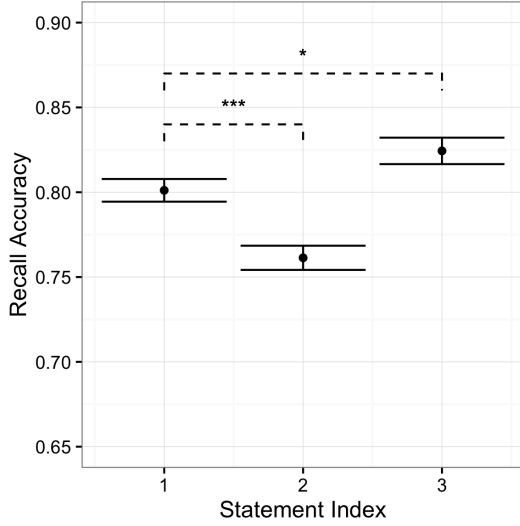


Figure 12. Mean recall accuracy and

standard errors for each statement position.

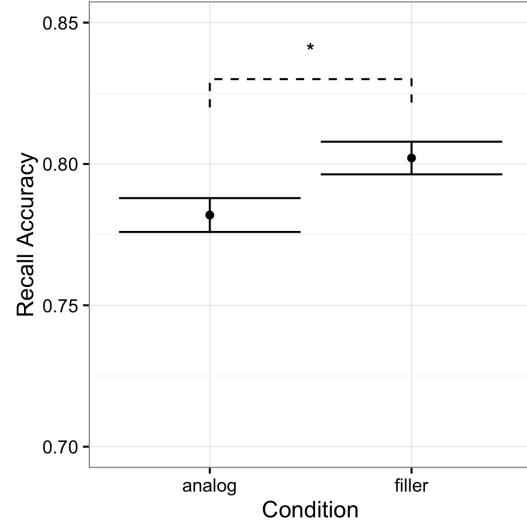


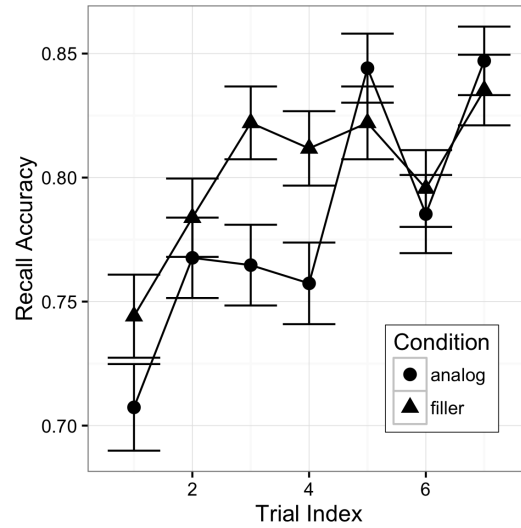
Figure 13. Mean recall accuracy and

standard errors for each story structure condition.

between the conditions. Nonetheless, due to the high statistical power the dummy variable comparing analog to filler items shows significance when analyzed with  $\chi^2$ -test ( $\chi^2(1, N = 9520) = 5.877, p = 0.015$ ).

**Main Effect.** Our main hypothesis was that repeated causal structure will lead to increased memory performance. Figure 14 depicts the interaction of story structure condition and trial index. At first sight, it looks quite messy and not like a consistent trend. But one has to consider, that in this visualization there's still a lot of noise embedded in the form of confounding effects. We tested the interaction on statistical significance with a crossed random effects logistic regression analysis predicting recall success by adapting a method from Baayen, Davidson, and Bates (2008). Table 1 shows an overview of the models tested with different confounding parameters. All models therein have a random intercept for each subject and each statement, accounting for random biases introduced by individuals and linguistic differences.

All models from table 1 are nested, which makes their log-likelihood comparable. The significance stars next to the log-likelihood indicates the significance level of the  $\chi^2$ -test comparing the respective model to the next simpler (the one to the left). To test the



*Figure 14.* Mean recall accuracy and standard errors for each story position and story structure condition.

interaction, we z-transformed the variables trial index and condition (story structure: analog vs. filler).

Model 1 is the simplest, only accounting for the single effects plus their interaction, which is our main focus. Bear in mind that those models are crossed random effects models with random subject and statement intercepts. We can see that trial index has a strong effect on recall success, which has already been visible in the figures 14 and 10. The interaction seems to be not too far off from significance but has not reached the 5%-cutoff. Although when adding additional controlling parameters (as in model 2, 3 and 4), the main interaction become more and more significant mainly due to its increased estimate. As suspected, the fit of the models increases when controlling for the confounding effects. According to the  $\chi^2$ -tests the loss of degrees of freedom seem justified. it is noteworthy that the increase in fit is quite large when adding the log of the time reading (model 4), compared to the increase in fit when adding "Statement Nr" (model 3).

Table 1

*Logistic Regression Models Predicting Recall Success*

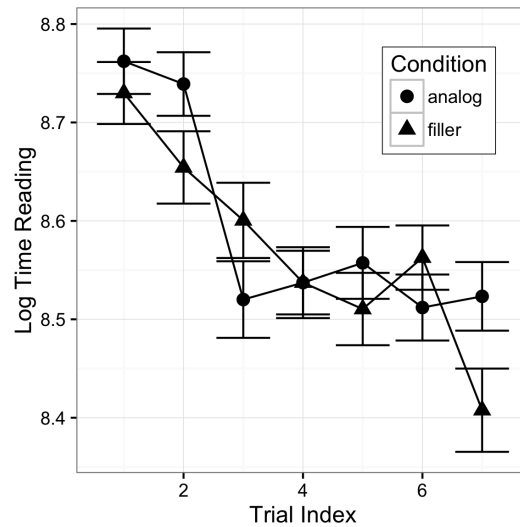
	<i>Dependent variable:</i>					
	Recall Success					
	(1)	(2)	(3)	(4)	(5)	(6)
Trial Index	0.218** (0.028)	0.223** (0.028)	0.222** (0.028)	0.280** (0.029)	0.327** (0.031)	0.256** (0.034)
Condition	0.071 (0.045)	0.073 (0.051)	0.072* (0.036)	0.076* (0.038)	0.071 <sup>o</sup> (0.038)	0.069 <sup>o</sup> (0.037)
Response Type Relation		-0.703** (0.057)	-0.724** (0.057)	-0.738** (0.058)	-0.740** (0.058)	-0.743** (0.058)
Statement Nr 2			-0.289** (0.083)	-0.022 (0.092)	-0.023 (0.090)	-0.023 (0.088)
Statement Nr 3			0.311** (0.093)	0.938** (0.113)	0.941** (0.112)	0.943** (0.110)
Log Reading Time				0.542** (0.048)	0.544** (0.048)	0.545** (0.048)
After Trick					-0.566** (0.112)	-0.535** (0.112)
First						-0.547** (0.110)
Trial Index:Condition	-0.052 <sup>o</sup> (0.028)	-0.053 <sup>o</sup> (0.028)	-0.055 <sup>o</sup> (0.028)	-0.057* (0.029)	-0.063* (0.029)	-0.057* (0.028)
Constant	1.726** (0.139)	2.036** (0.146)	2.051** (0.149)	-2.838** (0.449)	-2.810** (0.449)	-2.780** (0.449)
Observations	9,520	9,520	9,520	9,520	9,520	9,520
Log Likelihood	-4,191.375	-4,116.757***	-4,101.298***	-4,030.975***	-4,018.804***	-4,006.802***
Akaike Inf. Crit.	8,394.750	8,247.515	8,220.596	8,081.950	8,059.609	8,037.605
Bayesian Inf. Crit.	8,437.717	8,297.643	8,285.046	8,153.562	8,138.381	8,123.538

Note: <sup>o</sup> p<0.1; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001



To account for two anomalies in our data, we tested two additional models adding a free trial position parameter on each, the first parameter being "after trick". As we have seen in figure 10, recall success remarkably declines after the insertion of the trick story in our trial sequence. To investigate the effect of this incoherence on our main interaction, we ran the logistic regression in model 5 with an extra free parameter for all trials after the trick story. Not surprisingly the log-likelihood of this model increases significantly, and the coefficient estimate for interaction is bigger then in the previous models.

The other anomaly from inspecting the effect of trial index on recall accuracy (fig. 10) was the big deviation at the first trial. To investigate if the statistical power of the interaction mainly results from this deviation, we embedded another free parameter for all first trials. Even though we can observe a decreased coefficient estimate in the interaction in model 6, it stays below the 5%-significance level.



*Figure 15.* Mean and standard errors of the logarithm of the reading time in milliseconds for each trial index and story structure condition.

**Reading Time.** Day and Gentner (2007) observed decreased encoding time when participants read a story which was structurally similar to the previous story. Therefore we also tested the effect of repeated causal structure on encoding time with a crossed

random effects regression analysis, similar to when predicting recall success. This time it is not a logistic but a linear model. Also, the linear regression runs on the level of the statement (N=3570), and not item (N=9520). The interaction of story structure structure with trial index does not reach significance (see table 2). it is important to mention that our task crucially differed from Day and Gentner (2007) in the sense, that we specifically ask our participants to remember the different pieces of a very short story, while they have asked to actually comprehend a story with the expectation to answer questions about it. Also our material differs a lot in terms of complexity and conventionality.

Table 2

*Linear Regression Models Predicting the Log of Time Reading*

	<i>Dependent variable:</i>	
	Log Time Reading	
	(1)	(2)
Trial Index	−0.082** (0.010)	−0.082** (0.010)
Condition	−0.007 (0.070)	−0.007 (0.013)
Statement Nr 2		−0.422** (0.032)
Statement Nr 3		−1.064** (0.032)
Trial Index:Condition	−0.004 (0.010)	−0.004 (0.010)
Constant	8.511** (0.094)	9.007** (0.067)
Observations	3,570	3,570
Log Likelihood	−3,485.351	−3,421.937***
Akaike Inf. Crit.	6,984.703	6,861.874
Bayesian Inf. Crit.	7,027.965	6,917.497
<i>Note:</i>	° p<0.1; * p<0.05; ** p<0.01; *** p<0.001	

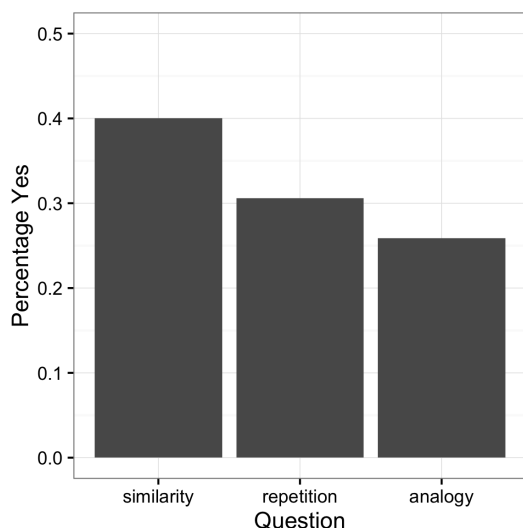


Figure 16. Percentage of positive responses to the three follow-up questions.

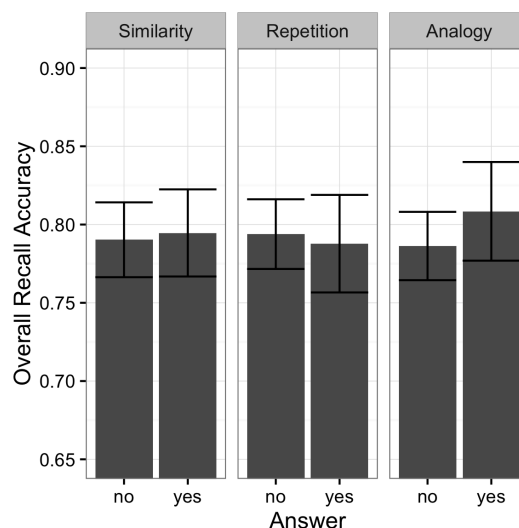


Figure 17. Mean and standard error of subjects overall recall accuracy for each response to the three follow-up questions.

**Follow-Up.** We claim that any analogical transfer in this experiment is unintentional, because participants are supposed to recall content and aren't told to explicitly compare stories to each other or even are aware of the systematicity. To check to what extent participants were aware of the repeated analog structure we asked three follow-up questions. The percentages of our sample that detected certain kinds of similarities between the stories are shown in figure 16. While 40% reported to notice similarities between the stories, when asked to explain their response, most participants mentioned that all stories had nonsensical names. Some also mentioned the linguistic coherence ("Every story had 3 names of places, things or people.", "Every story had 2 actions and 1 consequence"). 31% reported to having noticed repeating elements. The open answers were quite similar to the previous question. Only 26% of our sample said to having noticed the repeating analog/causal structure in every second story. Although their open responses reveal, that often they meant the syntactic similarity instead of the causal structure. Only a very few participants really detected the repeating analog structure, indicating that most people could not have intentionally mapped stories on to each other.

We further investigated if detection of certain kinds of similarities between the stories had an effect on subject's overall recall accuracy. Figure 17 depicts those comparisons, which seem to not make a difference. Running simple linear regressions also showed no significant effect for any of the questions (see Table 3), indicating that being aware of any kind of shared similarity did not yield any benefit in memory performance.

Table 3

	<i>Dependent variable:</i>		
	Overall Recall Accuracy		
	(1)	(2)	(3)
Similarity	0.004 (0.037)		
Repetition		-0.006 (0.039)	
Analogy			0.022 (0.041)
Constant	0.790** (0.023)	0.794** (0.022)	0.786** (0.021)
Observations	85	85	85
R <sup>2</sup>	0.0002	0.0003	0.003
Adjusted R <sup>2</sup>	-0.012	-0.012	-0.009
Residual Std. Error (df = 83)	0.167	0.167	0.167
F Statistic (df = 1; 83)	0.014	0.024	0.287
<i>Note:</i>	° p<0.1; * p<0.05; ** p<0.01; *** p<0.001		

## Conclusion

In this experiment we have discovered various weaknesses in the experimental setting. Nonetheless we could replicate a set of well known effects. Plus, there seems to be weak evidence for an increased learning rate with repeated causal structure. To strengthen our argument for analogical transfer in this hebb repetition paradigm, we decided to conduct another experiment, hoping to improve the methods where we see them unfit.

## Experiment 2

While in experiment 1 we only detected weak evidence for increased memory performance in repeated analogical structured stories, it could still be the case, that the effect is for example shadowed by the immediate change of story structure within the trial sequence. Opposed to relational priming studies, where the effect was usually observed when the same relation was activated just before the cued word pair. That's why we decided to investigate another experimental setting of immediate story structure repetition. Another surprise in experiment 1 was the overall high recall success rate. We observed that participants were taking a lot of time to encode and/or rehearse the information. A lot more than would occur in a natural reading environment. We therefore decided to change the free reading time to a fixed time. Due to the anomalies at trial position 1 and 11 in experiment 1, we also introduced a third category of stories, we called burn stories. These stories solely served the purpose of reducing noise from unintended trial sequence effects at the very beginning and after the trick story. They have the same characterization as the filler stories.

### Methods

The methods were the same as in Experiment 1 with following exceptions.

**Task.** The task differed only regarding the fixed reading time. For each statement participants had 3 seconds to encode the information, which is less than half of the median time reading from experiment 1.

**Participants.** We again collected observations from 78 participants on Crowdfunder. Applying the same exclusions criteria as in experiment 1, we ended up with a sample of 62 participants. Median age was 28 years, which is remarkably lower than in experiment 1 (see figure 19). Qualification was equally balanced between high-school and university degrees (each 50%). 60% percent of our sample was female.

**Material.** To further balance story difficulties we replaced a few outlier stories with new ones. Because of the additional need of filler stories, we constructed a few new ones. To see the full list of stories used in this experiment see the appendix.

Trial Sequences																	
S1:	B	B	B	A	A	F	F	A	A	F	F	T	B	A	A	F	F
S2:	B	B	B	F	F	A	A	F	F	A	A	T	B	F	F	A	A

Figure 18. Note: A=analog, B=burn, F=filler, T=trick

**Procedure.** In the preview description on CrowdFlower we changed the wording a little, because we didn't want participants to focus too much on effort. So we told them that they will get the bonus only if they pay attention to the task throughout the experiment, and that we will probe their attention. This way we also wanted to reduce the incentive to cheat. Performance should not play a role in payment. The biggest change in experiment 2 is the introduction of the repeated story structure condition plus a burn phase at the beginning and after the trick story. Figure 18 offers an overview of the two possible trial sequences in this experiment. Participants were assigned randomly to one of those timelines.

## Results and Discussion

We will only present and discuss results where they deviate from experiment 1.

**Material Assessment.** Even though we tried to assimilate the difficulties between the stories, it seems the reduced encoding time has introduced more sensitivity regarding story difficulty to the task (see figure 20). The standard deviation between the stories mean recall accuracies are bigger in experiment 2 than in experiment 1. Although the Levene test for homogeneity of variances is just marginally significant ( $F(1,24) = 3.082, p = 0.091$ ). Therefore the analysis of variance is again significant. But because we have less statistical power because we have less observations, this time the significance is on a lower level ( $F(11,5940) = 2.318, p = 0.008$ ).

**General Task Performance.** Setting a time limit during encoding clearly had a big impact on recall accuracy. Figure 21 depicts the distribution of participants' overall recall accuracy. Median is at 67% recall accuracy, more than 10% lower than in

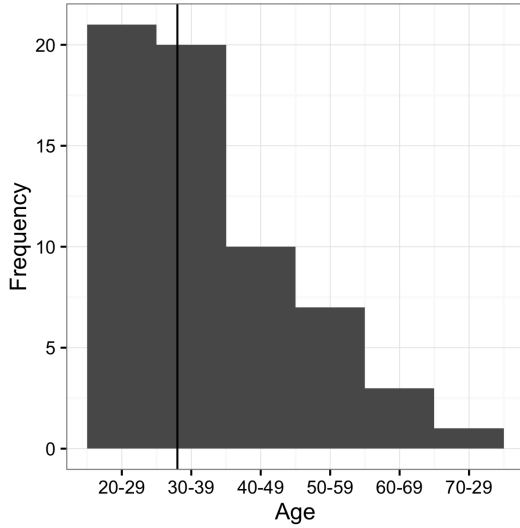


Figure 19. Distribution of age plus median in the sample of experiment 2.

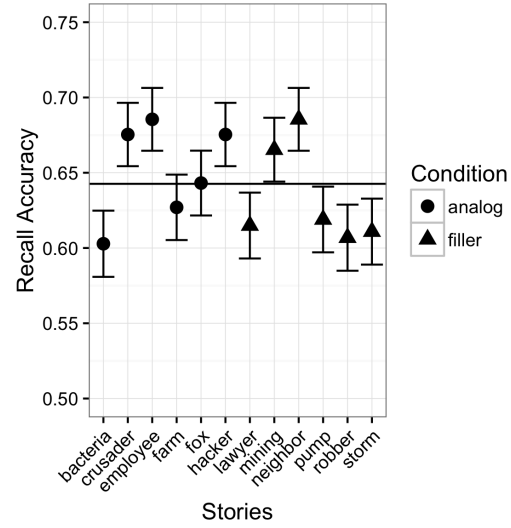


Figure 20. Mean and standard error of recall accuracy for each story with grand mean.

experiment 1. A Welch two sample t-test between subjects' overall recall accuracies from experiment 1 and 2 showed high significance ( $t(147)=4.74$ ,  $p < 0.001$ ). As a result from the increased difficulty, also the sensitivity to an individual's performance increased. The Levene test for homogeneity of variances showed a significant result ( $F(1,145) = 4.0749$ ,  $p = 0.045$ ).

**Confounding Effects.** In contrast to experiment 1 and to our surprise, we could not observe a increased memory performance effect over trial index in a simple logistic regression ( $e=0.001$ ,  $z(5950)=0.194$ ,  $p=0.846$ ). Figure 22 shows the absence of any tendency in recall accuracy over trial index. We can think of four reasons, why we could not replicate this finding from experiment 1. First (a), there is no such effect in this experimental setting and the evidence in experiment 1 was obtained by chance. Second (b), the increased difficulty through the limited encoding time made the task too difficult and less sensitive to such a subtle effect. Third (c), the introduced burn phase, the first three trials in the beginning of this second experiment, took the wind out of the sails, meaning there is only an increase in recall performance over the very first few trials. And fourth (d), the samples differed regarding age which could be responsible for

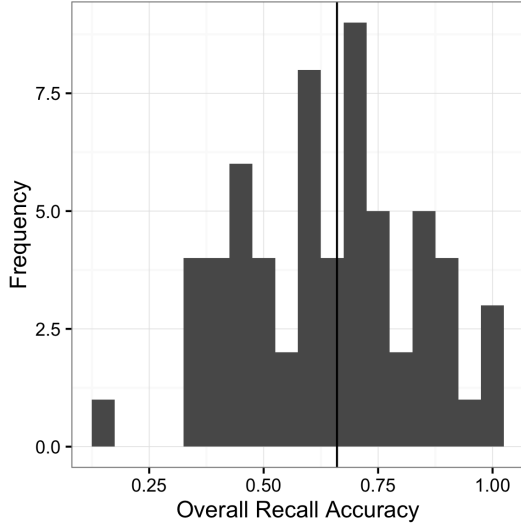


Figure 21. Distribution of subjects' overall recall accuracy with median.

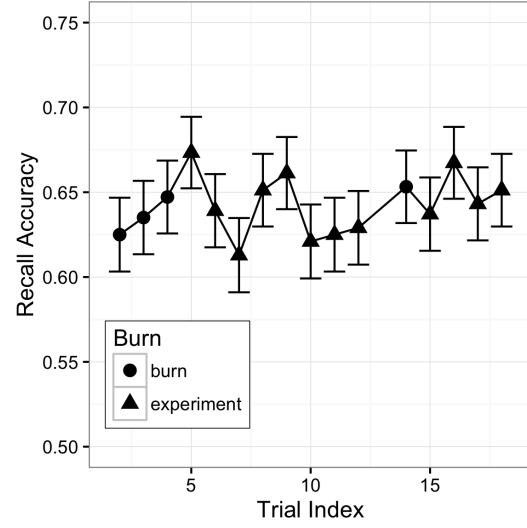


Figure 22. Mean and standard error of each trial position with grand mean.

the absence, as age and memory performance are known to be related (Hasher & Zacks, 1988).

The last hypothesis (d) is not supported by evidence. Even though a Welch's t-test between both sample's age distribution shows marginal significance ( $t(147) = 1.81$ ,  $p = 0.073$ ), when trying to predict subjects' overall recall accuracy with age while controlling for experiment differences, the predictor age does not play a significant role in this linear regression ( $e < 0.001$ ,  $t(144) = 0.182$ ,  $p = 0.856$ ).

The third hypothesis (c) is opposed by the fact, that even when excluding the first three trials in experiment 1, there's still a strong trial index effect on recall success in a logistic regression analysis ( $e = 0.029$ ,  $z(9518) = 3.558$ ,  $p < 0.001$ ), meaning that at least in experiment 1 the first three trials are not responsible for the trial index effect. Plus, this hypothesis is also contradictory to the logistic regression including the data from burn trials in experiment 2, which also does not show a trial position effect ( $e = 0.003$ ,  $z(7936) = 0.626$ ,  $p = 0.532$ ), which is also shown by figure 22.

The second hypothesis (b), that the increased difficulty is responsible for a reduced sensitivity for a trial index effect, is supported by the finding that we neither detected an "after trick" and nor a "first" effect in experiment 2. In experiment 1 we observed a



remarkable decrease in recall success after the trick story and greatly increased recall performance at the first trial. The logistic regression when controlling for trial index effects, neither shows significance for after trick trials ( $e=0.041$ ,  $z(7936) = 0.413$ ,  $p = 0.68$ ), nor the first trials ( $e = -0.065$ ,  $z(7936) = -0.624$ ,  $p = 0.533$ ). The concurrent absence of both effects indicate a reduced sensitivity of this experimental setting to learning effects, due to increased task difficulty.

Hypothesis one (a), that the results from experiment 1 were mere coincidence, can not be completely neglected, but given the alternative explanation for the absence of a trial index effect make it seem unlikely.

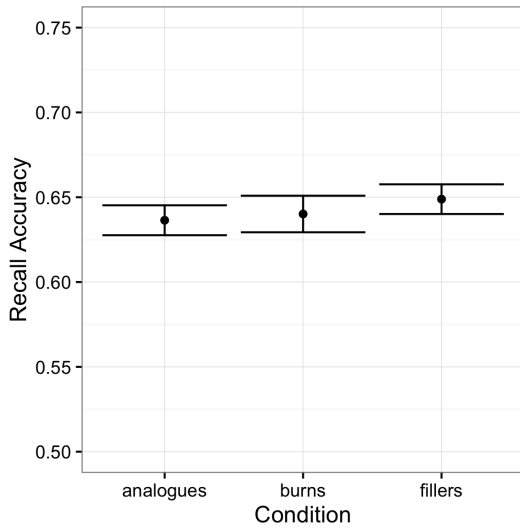


Figure 23. Mean and standard error for each condition.

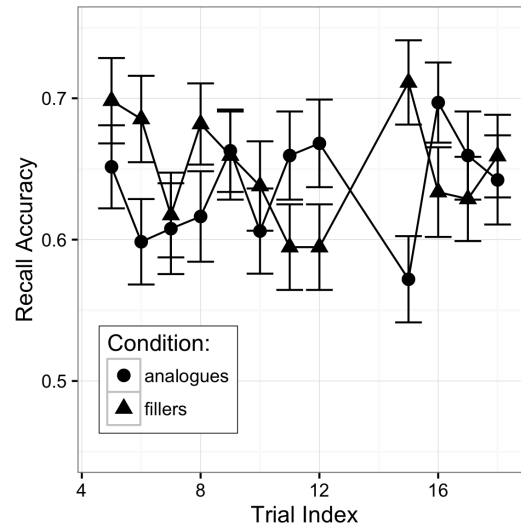


Figure 24. Mean and standard error for each trial index and condition .

The other confounding effects like statement positino and response type look very much like in experiment 1, although sometimes a little different accentuated.

**Main Effects.** Even though we can't observe a trial index effect, there still could be an interaction with the story structure condition. An improved recall performance in the analog condition could for example be shaded by stable performance throughout the experiment in the filler condition, which could have resulted in not finding a trial index effect. Also, with the immediate repetition of the conditions, we introduced a binary variable we called "primed" in experiment 2, reflecting if the previous trial was from the

Table 4

*Logistic Regression Models Predicting Recall Success for Experiment 2*

	<i>Dependent variable:</i>			
	Recall Success			
	(1)	(2)	(3)	(4)
Trial Index	0.009 (0.030)	0.008 (0.031)	0.008 (0.030)	0.008 (0.031)
Response Type Relation	−0.496** (0.063)	−0.496** (0.063)	−0.497** (0.063)	−0.497** (0.063)
Statement Nr 2	−0.323** (0.085)	−0.323** (0.083)	−0.323** (0.084)	−0.323** (0.084)
Statement Nr 3	0.692** (0.096)	0.692** (0.095)	0.693** (0.096)	0.693** (0.096)
Primed		−0.003 (0.061)		−0.003 (0.061)
Condition		0.026 (0.047)	0.033 (0.036)	0.030 (0.048)
Primed:Condition		0.013 (0.061)		
Trial Index:Condition			−0.035 (0.031)	−0.034 (0.031)
Condition:Primed				0.005 (0.061)
Constant	0.941** (0.167)	0.943** (0.170)	0.942** (0.167)	0.944** (0.170)
Observations	5,952	5,952	5,952	5,952
Log Likelihood	−3,321.002	−3,320.571	−3,319.984	−3,319.979
Akaike Inf. Crit.	6,656.004	6,661.143	6,657.968	6,661.958
Bayesian Inf. Crit.	6,702.844	6,728.058	6,718.191	6,735.564

*Note:*

° p&lt;0.1; \* p&lt;0.05; \*\* p&lt;0.01; \*\*\* p&lt;0.001

same condition or not. "Primed" could show additional benefits in recall through immediate repetition of a story structure, similar to priming studies. Because we expect this effect only to appear in the analog condition, we have to test the interaction while controlling for story index effects.

Table 4 shows an overview of calculated logistic crossed random effects models. Figure 24 shows the recall accuracy for each trial index and condition, which shows no clear tendency of an interaction between condition and trial index. This is confirmed in the model 3 and 4 of table 4 where the interaction shows no significance. The primed effect shows no significance in the models 2 and 4 of the same table, which can also be seen in figure 25. The standard errors are overlapping a lot.

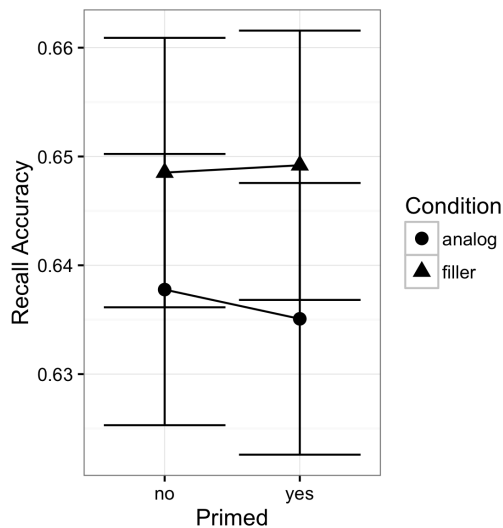


Figure 25. Mean and standard error for primed and condition.

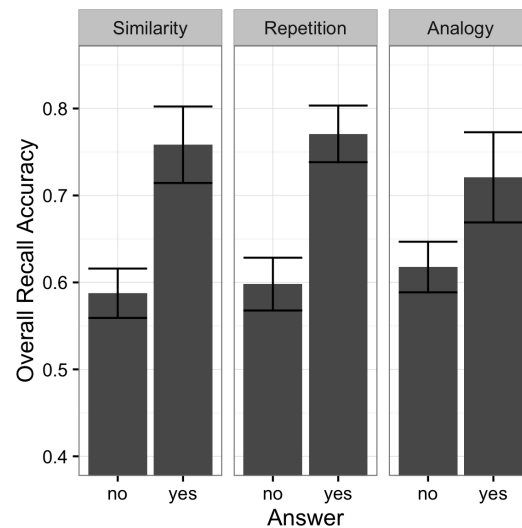


Figure 26. Mean and standard error of subjects' recall accuracy for each follow-up question.

We have already argued, that the increased task difficulty in the second experiment led to a decreased sensitivity for the trial index effect. The same could be true for even weaker effects like the interaction with condition or the effect of primed and condition. Although in this case, we have less evidence that speaks for this explanation. The alternative explanation, that those interaction effects were obtained by chance in the first experiment seems more likely, than when arguing against chance for the trial index

effect.

Table 5

	<i>Dependent variable:</i>		
	Overall Recall Accuracy		
	(1)	(2)	(3)
Similarity	0.171** (0.051)		
Repetition		0.173** (0.055)	
Analogy			0.103 <sup>o</sup> (0.059)
Constant	0.588** (0.029)	0.598** (0.028)	0.618** (0.029)
Observations	62	62	62
R <sup>2</sup>	0.157	0.141	0.048
Adjusted R <sup>2</sup>	0.143	0.127	0.032
Residual Std. Error (df = 60)	0.188	0.190	0.200
F Statistic (df = 1; 60)	11.177**	9.835**	3.030 <sup>o</sup>
<i>Note:</i>	<sup>o</sup> p<0.1; * p<0.05; ** p<0.01; *** p<0.001		

**Follow-Up.** The percentages of self-reported similarity detection between the stories were comparable to those of experiment 1. What differed though, was the effects of similarity detection on recall performance. Table 5 shows an overview over three linear models predicting overall recall performance, one for each of the three follow-up questions (see appendix). Figure 26 shows the differences in recall performance with respect to similarity detection. It appears to be the case, that people that were aware of the certain similarities or repeating elements had higher recall accuracies than participants that reported to not have noted any differences. We can only speculate if this effect appeared due to an increase in task difficulty.

**Conclusion.** In the second experiment we could not replicate most findings from the first experiment. Setting a limit of 3 seconds per statement to read, which is less than half of the median participants used to read the statements in experiment 1, changed how trial index relates to recall performance and the interaction with story structure

condition as well. We neither could observe the predicted interaction of primed structure stories with condition.

## General Discussion and Future Directions

### Material

Materials were constructed specifically for this experiment and not validated before using it in the experiments, neither regarding their structural nor surface similarity. We only tested different story lengths to find a suitable amount of statements per story. The argument therefore can be made, that the analog stories did not actually share structural similarity, an assumption we made when building those stories. Also, we claim that our stories generally share no or only very few surface features. Through independent ratings by pairwise comparisons of stories, the subjective perceptions of those different kinds of similarities could at least be investigated.

### Processing Depth

Another major concern regarding this experimental setup is: how well are participants integrating the information they were encoding? It is important to note this, because analog stories only share abstract information, a increased recall performance could only be observed, if the information is encoded on the required level of abstraction. The task they had to solve was pure recall and even though we are asking for relations between two entities, we did not check how well the information was integrated and how deep it was processed. Participants could just as well have remembered the sequence of items and not integrate the information in more abstract representations. Evidence that this might be important can be seen in a study from Spellman et al. (2001). They only detected a relational priming effect when they asked participants to pay attention and use the relation. Also it seems to be the case when comparing various relational priming studies, that the kind of task participants had to do, moderates the relational priming effect (Popov & Hristova, 2015). In one study participants were asked if a word pair makes sense (Estes & Jones, 2006), which requires a higher degree of information integration than for example in a lexical decision task. It could be said that our task resembles more a lexical decision task than a sensicality task in terms of information integration.

One possibility to investigate this problem in future experiments, could be to add the passively formulated relation as response options in the relation menu. Through this, it probably might not be possible to force deeper information integration during encoding, but at least it could be measured and therefore controlled, because people that only learn sequences of items and therefore integrate information less, will have more difficulties in responding to passively formulated relations. This could become visible in longer decoding times or lower recall success rates.

To incentivize deeper processing of the stories, one could also think of an additional task, like rating the sensicality of the conclusive statement, as was done in relational priming studies already (Estes & Jones, 2006; Popov & Hristova, 2015). This could also be done while varying the sensicality of those statements between subjects or not. The bacteria story for example ("Bacteria #1 emits enzyme #2", "Enzyme #2 dissolves the membrane of organ #3.") could have two differing conclusive statements, i.e. "Organ #3 fails." (original) and "Organ #3 rules" (new and nonsensical). This addition would also be interesting to investigate, because the task would become more comparable to previous research in relational priming, when measuring reaction times of the sensicality judgement as was done before.

Individual differences regarding processing depth have not been accounted for in this thesis. According to the Theory of Cognition some people are generally more motivated to elaborate than others as kind of a personality trait (Cacioppo & Petty, 1982; Cacioppo, Petty, Feinstein, & Jarvis, 1996). Therefore there also could be individual differences in our task regarding the abstraction of representations of the stories. Individual differences could be measured with either the original NFC-questionnaire (Cacioppo & Petty, 1982) or abbreviated forms of the NFC by Epstein, Pacini, Denes-Raj, and Heier (1996) like REI-questionnaires, which are part of the cognitive-experiential self-theory. Controlling for individual differences can reduce the noise further in a regression analysis and therefore strengthen the statistical power of a very subtle effect.

**Task difficulty**

As we have discussed already, we think that task difficulty moderates the effect of repeated story structure on recall success. The fact that more than one of the observed effects in experiment 1 are absent in experiment 2 speak for this hypothesis (interaction condition with trial index, first position, after trick position). To strengthen this argument a manipulation of the encoding time with for example 2, 4 and 8 seconds encoding time could collect further evidence in favor of this hypothesis. Also a manipulation of story length to vary task difficulty could be considered.



### Summary

The aim of this thesis was to investigate whether recall performance changes over repeated causal structures in analog stories, building on previous research from relational priming and analogical reminding studies. By combining an established method from memory research with original story material, we developed a novel experimental setting to investigate this hypothesis and contribute to growing evidence that analogical transfer happens also unintentionally. It would for the first time indicate that unintentional analogical transfer can also rely on mere structural similarity. We detected evidence for this effect in experiment 1, where we observed an interaction of story structure condition (analog vs. filler) and trial index on the the 5%-significance level. This effect could not be replicated in experiment 2, where we changed the trial sequence to a repeated condition setting and introduced limited reading time. Because various effects from experiment 1 concurrently could not be replicated in experiment 2, we argued that the most probable reason was the increased task difficulty through the limited encoding time. We then discussed further methodological improvements and possible future experimental endeavors.

## References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Memory and Language Mixed-effects modeling with crossed random effects for subjects and items. , *59*(4), 390–412. doi: 10.1016/j.jml.2007.12.005
- Cacioppo, J. T., & Petty, R. E. (1982). The Need of Cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131.
- Cacioppo, J. T., Petty, R. E., Feinstein, J. a., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253. doi: 10.1037/0033-2909.119.2.197
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What One Intelligence Test Measures: A Theoretical Account of the Processing in the Raven Progrogessive Matrices Test. *Psychological Review*, *97*(3), 404–431.
- Catrambone, R. (2002). The effects of surface and structural feature matches on the access of story analogs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 318–334. doi: 10.1037/0278-7393.28.2.318
- Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1992). High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental & Theoretical Artificial Intelligence*, *4*(3), 185–211. doi: 10.1080/09528139208953747
- Clement, A., & Gentner, D. (1991). Systematicity as a Selection Constraint in Analogical Mapping. *Cognitive Science*, *15*(1), 89–132.
- Cornelissen, J. P. (2003). Metaphor as a Method in the Domain of Marketing. *Psychology & Marketing*, *20*(2), 99–121. doi: 10.1002/mar. 10068
- Day, S. B., & Gentner, D. (2007). Nonintentional analogical inference. *Memory & Cognition*, *35*(1), 39–49.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*, 1–12. doi: 10.3758/s13428-014-0458-y

- Dunbar, K. N. (2001). The invivo/invitro approach to cognition: the case of analogy. *Trends in Cognitive Sciences*, 5(8).
- Dupin, J.-J., & Joshua, S. (1989). Analogies and “modeling analogies” in teaching: Some examples in basic electricity. *Science & Education*, 73(2), 207–224. doi: 10.1002/sce.3730730207
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, 71(2), 390–405. doi: 10.1037/0022-3514.71.2.390
- Estes, Z., & Jones, L. L. (2006). Priming via relational similarity: A copper horse is faster when seen through a glass eye. *Journal of Memory and Language*, 55, 89–101. doi: 10.1016/j.jml.2006.01.004
- Gagné, C. L., Spalding, T. L., & Ji, H. (2005). Re-examining evidence for the use of independent relational representations during conceptual combination. *Journal of Memory and Language*, 53(3), 445–455. doi: 10.1016/j.jml.2005.03.006
- Gentner, D. (1982). *Flowing Waters or Teeming Crowds: Mental Models of Electricity*. Cambridge: Bolt Beranek and Newman Inc.
- Gentner, D. (2002). Analogy in Scientific Discovery: The Case of Johannes Kepler. *Model-based Reasoning: Science, technology, values*, 21–39.
- Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy. *The American psychologist*, 52(1), 32–34. doi: 10.1037/0003-066X.52.1.32
- Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, 95(2), 393–405. doi: 10.1037/0022-0663.95.2.393
- Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The Roles of Similarity in Transfer: Separating Retrievability from Inferential Soundness. *Cognitive psychology*, 25(4), 524–575.
- Gick, M., & Holyak, J. K. (1980). Analogical Problem Solving. *Cognitive Psychology*, 12, 306–355. doi: 10.1016/0010-0285(80)90013-4

- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15(1), 1–38. doi: 10.1016/0010-0285(83)90002-6
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *The psychology of learning and motivation*, 22, 193–225. doi: 10.1016/s0079-7421(08)60041-9
- Herzenstein, M., & Hoeffler, S. (2014). Of clouds and zombies: How and when analogical learning improves evaluations of really new products. *Journal of Consumer Psychology*. doi: 10.1016/j.jcps.2016.01.002
- Hofstadter, D., & Sander, E. (2013). *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. New York, New York, USA: Basic Books.
- Hofstadter, D. R. (2001). Analogy as the Core of Cognition. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 499–538). doi: 10.1007/s13398-014-0173-7.2
- Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & cognition*, 15(4), 332–340. doi: 10.3758/BF03197035
- Hristova, P. (2009). Unintentional and Unconscious Analogies Between Superficially Dissimilar But. In *New frontiers in analogy research: Proceedings of the second international conference on analogy* (pp. 193–203).
- Johnson, H. M., & Seifert, C. M. (1992). The Role of Predictive Features in Retrieving Analogical Cases. *Journal of Memory and Language*, 31(5), 648–667.
- Kokinov, B. N. (1994). A hybrid model of reasoning by analogy. *Analogical Connections*, 2, 247–318.
- Lafond, D., Tremblay, S., & Parmentier, F. (2010). The ubiquitous nature of the Hebb repetition effect: error learning mistaken for the absence of sequence learning. *Journal of experimental psychology. Learning, memory, and cognition*, 36(2), 515–522. doi: 10.1037/a0018469
- Leech, R., Mareschal, D., & Cooper, R. P. (2008). Analogy as relational priming: a developmental and computational perspective on the origins of a complex cognitive skill. *The Behavioral and brain sciences*, 31(4), 357–378. doi:

10.1017/S0140525X08004469

- Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review*, 6(4), 586–597. doi: 10.3758/BF03212967
- Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, 7(4), 618–630. doi: 10.3758/BF03212999
- MacDonald, S. (2002). Hitler’s Shadow: Historical Analogies and the Iraqi Invasion of Kuwait. *Diplomacy & Statecraft*, 13(4), 29–59. doi: 10.1080/714000347
- Oberauer, K., Jones, T., & Lewandowsky, S. (2015). The Hebb repetition effect in simple and complex memory span. *Memory & cognition*, 43, 852–865. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/25712915> doi: 10.3758/s13421-015-0512-8
- Oesch, A. K. (2016). *Experiment: Analogical Transfer in a Hebb Repetition Paradigm*. Zenodo. Retrieved from <https://www.github.com/adrianoesch/analogical-transfer> doi: 10.5281/zenodo.59901
- Popov, V., & Hristova, P. (2015). Unintentional and efficient relational priming. *Memory & Cognition*, 43(6), 866–878. Retrieved from <http://link.springer.com/10.3758/s13421-015-0514-6> doi: 10.3758/s13421-015-0514-6
- Spellman, B. a., Holyoak, K. J., & Morrison, R. G. (2001). Analogical priming via semantic relations. *Memory & cognition*, 29(3), 383–93. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11407415> doi: 10.3758/BF03196389
- Treagust, D. F., Duit, R., Joslin, P., & Lindauer, I. (1992). Science teachers’ use of analogies: observations from classroom practice. *International Journal of Science Education*, 14(February 2015), 413–422. doi: 10.1080/0950069920140404
- Wharton, C. M., Holyoak, K. J., & Lange, T. E. (1996). Remote analogical reminding. *Memory and Cognition*, 24(5), 629–643. doi: 10.3758/BF03201088

## Appendix

### Material Experiment 1

**Instructions on CrowdFlower.** In this online experiment, participants will read a few, very short stories. After each story, participants are asked to recall the content of those stories by reconstructing its statements in the order of appearance. It will take around 20 minutes. There will be a confirmation code handed out at the end. We'll grant you a bonus of 2\$ if we can see that you showed effort. You must be a native English speaker and use a recent version of Chrome, Safari or Firefox to take part in this experiment. Means to facilitate the task are not allowed. You can only take part, if you weren't part of the same experiment already (check here). To start the experiment, please open the following link and carefully read the instructions.

### Follow-Up Questions.

1. Did you notice similarities between the stories during the experiment (not now)?
2. Did you notice repeating elements between the stories during the experiment?
3. Did you notice any repeating analog/causal structure during the experiment?

### Stories with Relation Response Options.

- Analog Stories:

Title	Statement	Relation-Options
Bacteria	Bacteria A emits enzyme B.	emits / expels / injects
	Enzyme B dissolves the membrane of organ C.	dissolves sth. of / disintegrates sth. of / disjoints sth. of
	Organ C fails.	fails / dies / is eliminated
Crusader	Crusader group A bought weapon B.	bought / acquired / caught
	Weapon B breaches the fortification of town C.	breaches sth. of / breaks sth. of / passes sth. of
	Town C is invaded.	is invaded / is intruded / is infiltrated
Hacker	Hacker A wrote computer code B.	wrote / invented / thought of
	Code B cracks the firewall of company C.	cracks sth. of / penetrates sth. of / catches sth. of
	Company C suffers a data loss.	suffers sth. / leaks sth. / tolerates sth.
Lawyer	Lawyer A finds loophole B.	finds / detects / meets

	Loophole B bypasses contract of union C.	bypasses sth. of / overrides sth. of / bystands sth. of
	Union C loses a lawsuit.	loses sth. / forfeits sth. / files sth.
Robber	Criminal A learns about security issue B. Security issue B suspends alarm system of museum C.	learns about / finds out about / starts looking suspends sth. of / disrupts sth. of / inspects sth. of
	Museum C gets robbed.	gets robbed / endures loss / is hijacked
Espionage	Country A trains spy B. Spy B is hired by Intelligence Agency of country C.	trains / instructs / creates is hired by sth. of / is employed by sth. of / is fired by sth. of
	Country C has its secrets stolen.	has sth. stolen / has sth. poached / has sth. lost
Fox	Fox A digs tunnel B. Tunnel B leads to coop of chicken C. Chicken C is devoured.	digs / burrow / wants leads to sth. of / gets to sth. of / goes on sth. of is devoured / is eaten / is missing

• Filler Stories:

Meteo	Cloud A collects humidity over ocean B.  Ocean B lies next to the beaches of area C. Area C is fertile.	collects humidity over / is contained in / lies in front of  lies next to / is located near / lies above is fertile / is humid / is rich
Fish	Fisher A cultivates sea grass B. Sea grass B is eaten by fish C. Fish C is captured.	cultivates / grows / breeds is eaten by / is consumed by / is accepted by is captured / is taken / is digested
Farm	Farmer A bought machine B. Machine B can collect fruit of plant C. Plant C are harvested automatically.	bought / purchased / borrowed can collect / garners / can compose are harvested automatically / are reaped mechanically / are picked manually
Mining	Mining company A has a lot of resource B. Resource B is sparse in country C. Country C starts negotiating.	has a lot of / is filled with / fills in for is sparse in / is rarely found in / is sparked through starts negotiating / starts brokering / starts trading
Insect	Perfume A attracts insect B. Insect B attacks woman C. Woman C has an allergic reaction.	attracts / lures / smells attacks / assaults / approaches has an allergic reaction / reacts allergically / has a panic reaction
Employee	Employee A is assigned to project B.	is assigned to / is referred to / is arranged to

	Project B supports agenda of manager C.	supports sth. of / emphasizes sth. of / destroys sth. of
	Manager C gets a promotion.	gets a promotion / gets an advancement / gets a bonus
School	School girl A hits boy B.	hits / punches / touches
	Boy B reports to teacher C.	reports to / tells / goes to
	Teacher C calls parents.	calls parents / notifies parents / visits parents

**Names.** List of randomly generated names for entities in stories. Names were generated using the free online tool [dialectcreator.com](http://dialectcreator.com):

Titho / Ilres / Fozeh / Thygef / Dohod / Rove / Eshbu / Ottha / Thuha / Epnux /  
 Evib / Potho / Ribi / Kawud / Avyth / Wehif / Uyim / Saly / Ohuy / Ezol / Eshov /  
 Lyde / Huluw / Gena / Lenik / Upum / Hoyl / Afi / Yuxeh / Thisun / Ofes / Matmag  
 / Vuloz / Puvi / Orok / Pishi / Thyra / Diaf / Hosen / Upno / Kewos / Hoap / Adjo /  
 Olag / Uwit / Ebib / Rivo / Uhuy / Ihab / Vivaf / Felac / Newi / Ahow / Mabi /  
 Owih / Rothuk / Napib / Assit / Eyem / Roze / Awiz / Zimob / Ifeh / Zidep / Wutoc  
 / Molash / Ofpoz / Mawi / Afik / Reok / Yoges / Uvman / Vilux / Lothe / Inag /  
 Goxos / Itbob / Mewep / Ewey / Meuf / Owmi / Upel / Eefar / Igoth / Olef / Ubug /  
 Aid / Hadith / Enmob / Wecin / Turow / Paza / Hize / Gelyp / Idog / Vipal / Pide /  
 Gabit / Zishu / Henu / Soav / Misug / Kesal / Dugoz / Pobat / Thusoth / Ohan /  
 Romap / Zugew / Ena / Niem

## Material Experiment 2

**Instructions on CrowdFlower.** In this online experiment, participants will be asked to read very short stories and recall its content after each story. The task is hard, don't expect to remember everything, but put in effort nonetheless. We will probe your attention. It will take around 15 to 20 minutes. There will be a confirmation code handed out at the end, but only if you pay attention to the task. You must be a native English speaker and use a recent version of Chrome, Safari or Firefox to take part in this experiment. We'll hand out a 1\$ bonus if you complete and pay attention. Also you can



only take part if you weren't part of another experiment from us (check here). To start the experiment, please open the following link and carefully read the instructions: [Link](#)

**Stories with Relation Response Options.**

- Analog Stories:

Title	Statement	Relation-Options
Bacteria	Bacteria A emits enzyme B.	emits / expels / injects
	Enzyme B dissolves the membrane of organ C.	dissolves sth. of / disintegrates sth. of / disjoints sth. of
	Organ C fails.	fails / dies / is eliminated
Crusader	Crusader group A buys weapon B.	buys / acquires / catches
	Weapon B breaches the fortification of town C.	breaches sth. of / breaks sth. of / passes sth. of
	Town C is invaded.	is invaded / is intruded / is infiltrated
Hacker	Hacker A writes computer code B.	writes / invents / thinks of
	Code B cracks the firewall of company C.	cracks sth. of / penetrates sth. of / catches sth. of
	Company C suffers a data loss.	suffers sth. / leaks sth. / tolerates sth.
Lawyer	Lawyer A finds loophole B.	finds / detects / meets
	Loophole B bypasses contract of union C.	bypasses sth. of / overrides sth. of / bystands sth. of
	Union C loses a lawsuit.	loses sth. / forfeits sth. / files sth.
Robber	Criminal A learns about security issue B.	learns about / finds out about / starts looking
	Security issue B suspends alarm system of museum C.	suspends sth. of / disrupts sth. of / inspects sth. of
	Museum C gets robbed.	gets robbed / endures loss / is hijacked
Fox	Fox A digs tunnel B.	digs / burrow / wants
	Tunnel B leads to coop of chicken C.	leads to sth. of / gets to sth. of / goes on sth. of
	Chicken C is devoured.	is devoured / is eaten / is missing

- Filler Stories:

Storm	Storm A hits tree B.	hits / blows off / runs over
	Tree B falls on trail of route C.	falls on / drops on / is thrown on
	Route C is blocked.	is blocked / is jammed / is ruined
Farm	Farmer A purchases machine B.	purchases / buys / borrows
	Machine B can collect fruit of plant C.	can collect / garners / can compose
	Plant C are harvested automatically.	are harvested automatically / are reaped mechanically / are picked manually

Mining	<p>Mining company A has a lot of resource B.</p> <p>Resource B is sparse in country C.</p> <p>Country C starts negotiating.</p>	<p>has a lot of / is filled with / fills in for</p> <p>is sparse in / is rarely found in / is sparked through</p> <p>starts negotiating / starts brokering / starts trading</p>
Pump	<p>Plumber A installs pump B.</p> <p>Pump B sucks water from ground C.</p> <p>Ground C dries up.</p>	<p>installs / builds / fixes</p> <p>sucks sth. from / transports sth. from / elevates sth. from</p> <p>dries up / dehydrates / soaks up</p>
Employee	<p>Employee A is assigned to project B.</p> <p>Project B supports agenda of manager C.</p> <p>Manager C gets a promotion.</p>	<p>is assigned to / is referred to / is arranged to</p> <p>supports sth. of / emphasizes sth. of / destroys sth. of</p> <p>gets a promotion / gets an advancement / gets a bonus</p>
Neighbor	<p>Man A takes care of cat B.</p> <p>Cat B walks over the flowers of neighbor C</p> <p>Neighbor C gets angry.</p>	<p>takes care of / looks after / is feeding</p> <p>walks over sth. of / goes over sth. of / strays over to</p> <p>gets angry / becomes furious / gets aroused</p>

- Burn Stories:

Fish	<p>Fisher A cultivates sea grass B.</p> <p>Sea grass B is eaten by fish C.</p> <p>Fish C is captured.</p>	<p>cultivates / grows / breeds</p> <p>is eaten by / is consumed by / is accepted by</p> <p>is captured / is taken / is digested</p>
School	<p>School girl A beats boy B.</p> <p>Boy B reports to teacher C.</p> <p>Teacher C calls parents.</p>	<p>beats / punches / touches</p> <p>reports to / tells / goes to</p> <p>calls parents / notifies parents / visits parents</p>
Insect	<p>Perfume A attracts insect B.</p> <p>Insect B attacks woman C.</p> <p>Woman C has an allergic reaction.</p>	<p>attracts / lures / smells</p> <p>attacks / assaults / approaches</p> <p>has an allergic reaction / needs medical support / has a panic reaction</p>
Meteo	<p>Cloud A collects humidity over ocean B.</p> <p>Ocean B lies next to the beaches of area C.</p> <p>Area C is fertile.</p>	<p>collects humidity over / is contained in / lies in front of</p> <p>lies next to / is located near / lies above</p> <p>is fertile / is humid / is rich</p>



## **Selbstständigkeitserklärung**

Hiermit erkläre ich, dass  
die Masterarbeit von mir selbst und ohne unerlaubte Beihilfe verfasst worden ist und ich die  
Grundsätze wissenschaftlicher Redlichkeit einhalte (vgl. dazu:  
[http://www.lehre.uzh.ch/plagiate/20110314\\_LK\\_Merkblatt\\_Plagiat.pdf](http://www.lehre.uzh.ch/plagiate/20110314_LK_Merkblatt_Plagiat.pdf)).

.....  
Ort und Datum

.....  
Unterschrift