



BACHELOR THESIS

Prototype or Exemplar: Can semantic representations in the brain predict word categorization?

Conceptual Learning

Pia Laura Elsasser

s1986619

p.l.elsasser@student.utwente.nl

First Supervisor: Prof. Dr. F. van der Velde

Second Supervisor: M.W. Westerhof, Msc

University of Twente

Faculty of Behavioural, Management and Social sciences (BMS)

Cognitive Psychology and Ergonomics (CPE)

July, 2020

UNIVERSITY OF TWENTE.

Table of Contents

Abstract	3
1. Introduction	4
1.1 From classical theory to prototypes and exemplars.....	5
1.2 Semantic representation as analysed by Huth et al. (2016)	7
1.3 Aim of the study	8
2. Method	10
2.1 Participants	10
2.2 Materials	10
2.3 Procedure	11
2.4 Data Analysis.....	11
3. Results	12
3.1 Card sorting task.....	12
3.2 Questionnaire.....	15
3.2.1 Mean ratings for category items.....	15
3.2.2 Mean ratings for filler items.....	19
4. Discussion	20
4.1 Limitations and recommendations for further study	22
4.2 Conclusion	23
5. References	24
6. Appendices	27

Abstract

Categorizing and relating concepts play a crucial role in the way we view and interact with the world. They give insights into the human mind and are essential in human reasoning. Regarding categorization, exemplar and prototype theory have remained prominent views. However, it is not clear which theory applies best when words are categorized. As semantic memory forms the basis of categorization, this thesis took the perspective of brain activation and, more specifically, semantic representations in the brain, to find out if exemplars or prototypes apply. For that, findings by Huth, de Heer, Griffiths, Theunissen, and Gallant (2016) were taken for further study. In their experiment, 11 word categories were found based on brain activation. The current research used items from these categories in a card sorting task to compare how participants would group them manually. Results showed that overall semantic representations in the brain are not able to predict manual word categorization. Further, participants created small and specific groups which indicates that word categorization is based on exemplars and not prototypes.

Keywords: categorization, exemplar theory, prototype theory, semantic memory, brain activation, card sorting

1. Introduction

Every day we use language to communicate with the people around us, we give meaning to the words we receive and understand what they are supposed to represent together formed as a sentence. Basis for the ability to give meaning to words and events in speech is our semantic memory. Tulving (1972) defines semantic memory as an accumulated knowledge a person has. This knowledge concerns word meanings or understanding of verbal symbols and, more specifically, rules and formulas for manipulating relations between them. Further, semantic memory includes general knowledge about facts, concepts, as well as objects, events and their properties and behaviours (Jones, Willits, & Dennis, 2015; Patterson, Nestor, & Rogers, 2007).

Fundamental to this is categorization, which helps in using the concepts in semantic memory and organizing our knowledge (Grossman et al., 2002). Categorizing and relating concepts play a crucial role in the way we view and interact with the world. They give insights into the human mind and are essential in human reasoning (Cai, Au Yeung, & Leung, 2012).

The current thesis will investigate theories surrounding categorization as there have been changes over time in what is believed to be the best fitting view. Focus will be on the differences between prototype and exemplar theory. The theories present a response to the classical view which was leading till the 1970s. This view states that categories have clear boundaries so that concepts belong to one group based on their attributes (Medin & Smith, 1984). In contrast to that, prototype and exemplar theory are based on similarity comparison. This means that for categorization people compare new concepts to an instance they have stored in their brain. They do not categorize solely based on features of the concept. In prototype theory the new concept is compared to a stored summary representation of the respective category (Rosch, 1983). In difference to that, exemplar theory holds that a new concept is compared to a specific example that the person has encountered before (Nosofsky & Zaki 2002). Both theories remained relevant but present different ideas and approaches to categorization.

This thesis will address this difference from the perspective of brain activation and, more specifically, semantic representations in the brain. For that, work by Huth, de Heer, Griffiths, Theunissen, and Gallant (2016) will be highlighted. Here, semantic representation was mapped across the cortex by analysing brain activity of seven subjects listening to narrative stories. 11 large word categories were found. The current research will build on Huth et al.'s results to find out whether exemplars or prototypes apply best. This will be done

by testing if the found semantic representations are able to predict manual word categorization in a card sorting task. This entails making groups from a set of words taken from categories identified by Huth et al. (2016). This task was chosen to assess the mental models of the participants. Word groups resulting from the task will be compared to the initial groups taken from Huth et al. and will further give insight into participants categorization style. Supplementary, a questionnaire will be administered in which Huth et al.'s found semantic representations are rated.

The remainder of this thesis is structured as follows: section 1.1 will discuss the development from classical view to prototype and exemplar theory and describe how they differ from each other. Section 1.2 discusses the work by Huth et al. (2016) in more detail and section 1.3 describes the aim of the current research. Moreover, the method and results can be found under section 2 and 3, respectively. Lastly, section 4 presents a discussion of the results and includes recommendations for further study.

1.1 From classical theory to prototypes and exemplars

As mentioned before, research on categorization and concept representation has a long history and with that views changed over the years. The general notion till the 1970s was that concepts are defined by specific features or attributes, which determine a clear category membership (Medin & Smith, 1984; Smith & Medin, 1981). This view is known as the classical view or classical theory. Here, categories are thought to be mutually exclusive with precise boundaries meaning that a concept needs to have all necessary attributes to belong in it. Consequently, a concept can only belong to one category (Cai et al., 2012; Medin & Smith, 1984).

However, classical theory turned out to be an incomplete explanation regarding categorization and thus was criticised greatly. Main problems concerned the specification of said concept attributes, category boundaries and the typicality effect (Cai et al., 2012; Medin & Smith, 1984). In different experiments it was found that people are not able to produce a list of necessary defining attributes and can be confused as to what category fits a certain concept. People often disagree between each other and even give inconsistent answers when asked at different times (Barsalou, 1989; Bellezza, 1984; McCloskey & Glucksberg, 1978). Additionally, the classical view cannot explain the typicality effect. The APA Dictionary of Psychology defines this effect as a preference for typical category members over atypical ones. People find it easier to make judgements about concepts that “represent” the category best. For instance, they are quicker to say that a dog is a mammal than they are to say that a

whale is a mammal (“Typicality effect”, 2020). In contrast to that, classical categorization assumes equal status for each category member because the concepts that belong together all share the same properties (Cai et al., 2012; Medin & Smith, 1984). Because of these problems arising, new improved theories had to follow.

One popular view is the prototype theory proposed by Rosch and colleagues (1976, 1983). Instead of using features to define concepts, the theory focuses on organizing concepts around resemblance to a category member that represents its group best. This member serves as a prototype, a summary representation, of the category (Aerts, Broekaert, Gabora, & Sozzo, 2016; Cruse, 2001; Nosofsky & Zaki 2002). Therefore, it must possess a wide range of features linked to the category. As concepts only relate to the centred prototype, and not match perfectly, it follows that category memberships have different degrees of fit. Not all members have equal status. Further, category boundaries are not clear-cut but rather vague and fuzzy (Cruse, 2001). This idea of fuzzy sets also connects to, for example, the results in the work of McCloskey and Glucksberg (1978) in which the typicality of category members was tested.

A second central view to categorization is the exemplar theory, a modification of the prototype theory (Cai et al., 2012). According to this view, categories are represented by stored members examples that belong to the specific category (Nosofsky & Zaki 2002). These members have been encountered by the person before and are called exemplars. When classifying new instances, the item is compared to the exemplar from the person’s memory (Nosofsky, 2011). If it is similar enough it will be classified as belonging to the same category the exemplar comes from.

When looking at both prototype and exemplar theory, it becomes apparent that they are based on similarity instead of rules as found in the classical view (Aerts et al., 2016). Instances are categorized by comparison and not by what attributes they possess. The two theories are alike in the way they describe the process of categorization but differ in their idea of what concepts are compared to during that process: a summary representation or specific encountered examples. Because of this, there have been opposing research findings and opinions.

For instance, work by Minda and Smith (2001) demonstrated the importance of prototypes. In four experiments they showed that a prototype model fit better than an exemplar one. More specifically, this effect was highlighted when participants learned large categories or contained complex stimuli. Murphy (2016), who reviewed past research, proposes that exemplar theory of concepts does not exist in a broad sense. Reason for this is

that different phenomena of concepts, such as hierarchical structures in categories for example, have received either none or incomplete exemplar explanations when for each phenomenon a prototype model was proposed or can be devised easily.

In contrast to that, research by Voorspoels, Vanpaemel, and Storms (2008) compared an exemplar model and a prototype model to find out how natural language categories are represented. Participants rated the typicality of various items in relation to their category. The results showed that an exemplar model fits better, which opposes that categories are represented by prototypes or summary representations. Moreover, in experiments with simple perceptual figures participants favoured the exemplar theory (Dopkins & Gleason, 1997). Rouder and Ratcliff (2006), who compared exemplar- and rule-based theories in connection to visual stimuli categorization, speculate that for complex stimuli such as words, exemplar theory will fit best.

Summarizing these findings, it is not clear which theory applies best for word categorizing. When learning categories, prototypes were found most applicable, but in connection to typicality and perceptual figures exemplars served best.

1.2 Semantic representation as analysed by Huth et al. (2016)

Since semantic memory and semantic information form the basis of categorization, they are important aspects to consider. Because of that connection, research on the location of semantic information in the brain might give insights into the nature of semantic memory. This can further reveal something about how words are categorized. For example, Grossman et al. (2002) analysed the neural basis of categorization and found that large-scale neural networks are associated with it in semantic memory.

A recent study which has focused on semantic representation was done by Huth, de Heer, Griffiths, Theunissen, and Gallant (2016). The researchers' motivation was to map the semantic system and analyse semantic selectivity of different brain regions, as this has not been done comprehensively before. In this study seven participants listened to ten narrative stories taken from "The Moth Radio Hour". Each story was about 10-15 minutes long. While the subjects listened, functional magnetic resonance imaging (fMRI) recorded their whole-brain blood-oxygen level dependent (BOLD) responses. Furthermore, to estimate semantic selectivity, voxel-wise modelling (VM) was applied as it is highly effective when modelling responses to complex natural stimuli (Huth, Nishimoto, Vu, & Gallant, 2012). A voxel (created from the words "volume" and "pixel") is a 3-dimensional unit (Torre, 2017). The brain can be divided into these voxels and for each brain activity can be measured by fMRI.

Concerning the results of this study, it was found that semantically selective areas are relatively symmetrically distributed across the two cerebral hemispheres. Additionally, the areas respond to different word clusters (Figure 1). In order to identify and label these clusters, the researchers constructed a 10,470 word lexicon from all words appearing in the stories. Eventually, 11 categories were found which were labelled “tactile”, “visual”, “number”, “outdoor”, “body part”, “place”, “violence”, “person”, “mental”, “time”, and “social”.

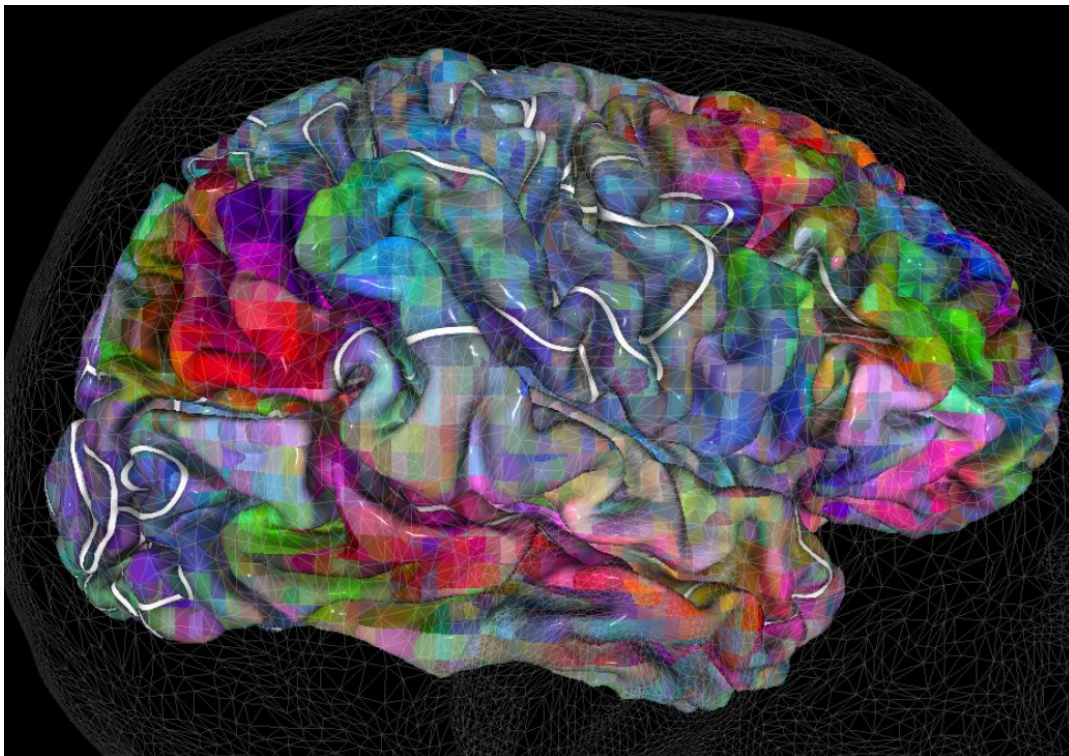


Figure 1. Semantic map (Gallantlab.org, n.d). Semantic selectivity is mapped across cortical surface based on Huth et al. (2016). Different colours indicate which word category is predicted to generate brain activity at each voxel. When clicking on a voxel, more information can be obtained on which specific words elicit brain activity.

1.3 Aim of the study

The prior literature assessment on prototype and exemplar theory points into different directions, especially as the mentioned findings are based on dissimilar experiments. However, Rouder, and Ratcliff (2006) speculate that exemplars are more fitting when categorizing words. This speculation needs to be confirmed or denied.

Further, Huth and colleagues' (2016) work on semantic representation found eleven large categories in 10,470 words based on blood oxygen levels in participants' brains. Based

on the size of these groups, prototypes might be more applicable, as explained by Minda and Smith (2001), if the found semantic representations are able to predict categorization in participants. This, of course, needs to be tested.

Based on this, the current thesis poses the following research questions:

I. *“Can semantic representations in the brain predict word categorization?”*

II. *“Is word categorization based on prototypes or exemplars?”*

In order to find answers, the present study will build on Huth et al.’s (2016) research by testing if the categories they found also apply when the categorization is done manually by subjects. It will be analysed if their categorization is based on exemplars or prototypes. For this, a card sorting task will be given. Card sorting tasks are used to elicit knowledge on a domain, more specifically the task produces representations of concepts and their interrelations (Cooke, 1994). So, this task will give info on participants’ mental models and will show in what way they categorize words. The task includes a set of words taken from Huth et al.’s brain map which is to be sorted into categories. This is done in three rounds, meaning that after the person created the categories, they are asked to further divide these into smaller groups.

As opposed to the fMRI measures of brain activity in Huth et al. (2016), the categories created in a card sorting task are based on the judgement of the participants. Because of the fundamental differences in these experiments, card sorting might be able to reveal aspects of categorization that were not found by Huth et al. (2016). Therefore, giving insights into the relation between brain mapping and categorization.

The card sorting task is followed by a questionnaire in which the fit between words and their assigned categories is rated. Participants indicate how similar they think the items are to Huth et al.’s (2016) categories. The aim of the questionnaire is to see if subjects validate Huth et al.’s found item-category word pairs.

Lastly, as the current research is conducted during the coronavirus pandemic, an online version of the experiment is included. Here, the card sorting task is reduced to one round so that further division of categories is left out.

2. Method

2.1 Participants

There were 20 participants with 13 females and 7 males and an average age of 20.9 years ($SD = 2.09$ years). Regarding the highest level of education, 18 participants had a high school diploma (or equivalent) and 2 were in the possession of a bachelor's degree. The participants were collected through SONA and by convenience sampling within the researcher's informal network. A sufficient level of English was required to be able to participate. Regarding this study, ethical approval was obtained by the BMS Ethics Committee of the University of Twente. As a reward, course credits were given to the participants who were in possession of a SONA account. Other participants did not receive any other form of reward. Each participant gave their informed consent (Appendix A) prior to the study.

2.2 Materials

For this study, a set of 50 paper cards was used. Each card had a word written on it. Words were taken from the brain map by Huth et al. (2016) (Figure 1) based on voxels that belonged to the category of either "time", "person", "place", "number" or "tactile" (Appendix B). From each of the five categories 10 words were taken so that each category was equally represented based on number of words. Furthermore, requirements were that the items are taken from two different voxels with at least a model performance of "not bad, pretty reliable". Model performance describes how well a voxel responds to the word categories. So, voxels with a low performance were not chosen as they are not reliable in their selectivity. Additionally, the chosen voxels were distributed over both hemispheres. An example item would be the word "fluid" belonging to category "tactile" which was taken from a voxel in the left hemisphere.

Secondly, a questionnaire was utilized (Appendix C) which consisted of 70 items measured with a 5-point Likert scale. In each question the participant rated a pair of words on how well they fit together. Choices ranged from "highly related" to "highly not related". Pairs were made from the 50 chosen words plus 20 filler words (Appendix B) and the five category names. Filler words were added to control for a response bias. Further, filler words had the same requirements for being chosen, however they were not allowed to be from one of the five categories mentioned before. For example, a pair to rate was "dollar-number" and filler pair was "cruelty-number". The questions were randomized so that a biased response from the participants could be further avoided.

2.3 Procedure

Participants were seated at a table and received general instructions of the study. After that they had to sign an informed consent. The study started with handing out the 50 word cards. Participants had to sort these into groups however they liked or it made sense to them. No naming of the groups was required. When the participant gave a sign that they were finished a picture was taken of each card group. Hereafter, in a second and third round, the participant had the choice to subdivide the existing groups. Again, pictures were taken of the new groups (if they were made).

The next part concerned filling in the questionnaire. This was done on a computer. The first block of questions concerned demographics such as age, gender, and highest level of education. The second block included rating the 70 word pairs. The experiment lasted about 30-40 minutes.

Online Version

Due to the regulations made during the time of the coronavirus pandemic, the initial experiment design was translated into an online version. The only change made concerned the card sorting task which was reduced to one round instead of a maximum of three. The experiment was created in Qualtrics and distributed via an anonymous link. 18 of the 20 participants took part in the online version. The other two participants took part in the regular experiment in a face-to-face setting.

2.4 Data Analysis

Before the analysis of the results of the card sorting task, the Jaccard score for each participant had to be calculated. This score represents the relation between two items and is constructed by counting the number of groups both items are member of divided by the number of groups at least one item belongs to (Schmettow & Sommer, 2016). To represent the scores a table was created in which the 50 words are both the rows and the columns.

In order to obtain the overall results of the card sorting task all Jaccard score tables were added which gave an unorganized heatmap. After that, a vector analysis of clusters was executed. This analysis was chosen as it aligns with Huth et al.'s (2016) work in which concepts were also represented as vectors. The R script used for this analysis can be found in Appendix D.

Furthermore, from the questionnaire the mean fit between items and category was calculated to see if participants validate the item-category pairs taken from Huth et al. (2016).

For this, the scale was scored as follows: 0 = highly not related, 1 = not related, 2 = neutral, 3 = related, 4 = highly related.

3. Results

3.1 Card sorting task

Starting with the vector analysis of the clusters, results were represented in an ordered heatmap (Figure 2). This represents the categorization done by all subjects. As indicated by the color key, red or “warm” areas show a high similarity between items whereas light blue shows “cold” areas where items have low similarity. At the diagonal clusters or groups of words can be identified. Here, 5 clusters were taken as the card sorting task was based on 5 initial categories identified by Huth et al. (2016). For comparison, words per heatmap cluster and per initial categories are listed in Table 1 and Table 2, respectively. Outside of that, yellow circles in the heatmap (1-4) show ambiguity of words. They highlight words that not only show similarities to words of their own cluster but also from other clusters. Meaning that participants somewhat differed in their choices of which words belong together. Therefore, the circles are located outside of the diagonal cluster row as outliers.

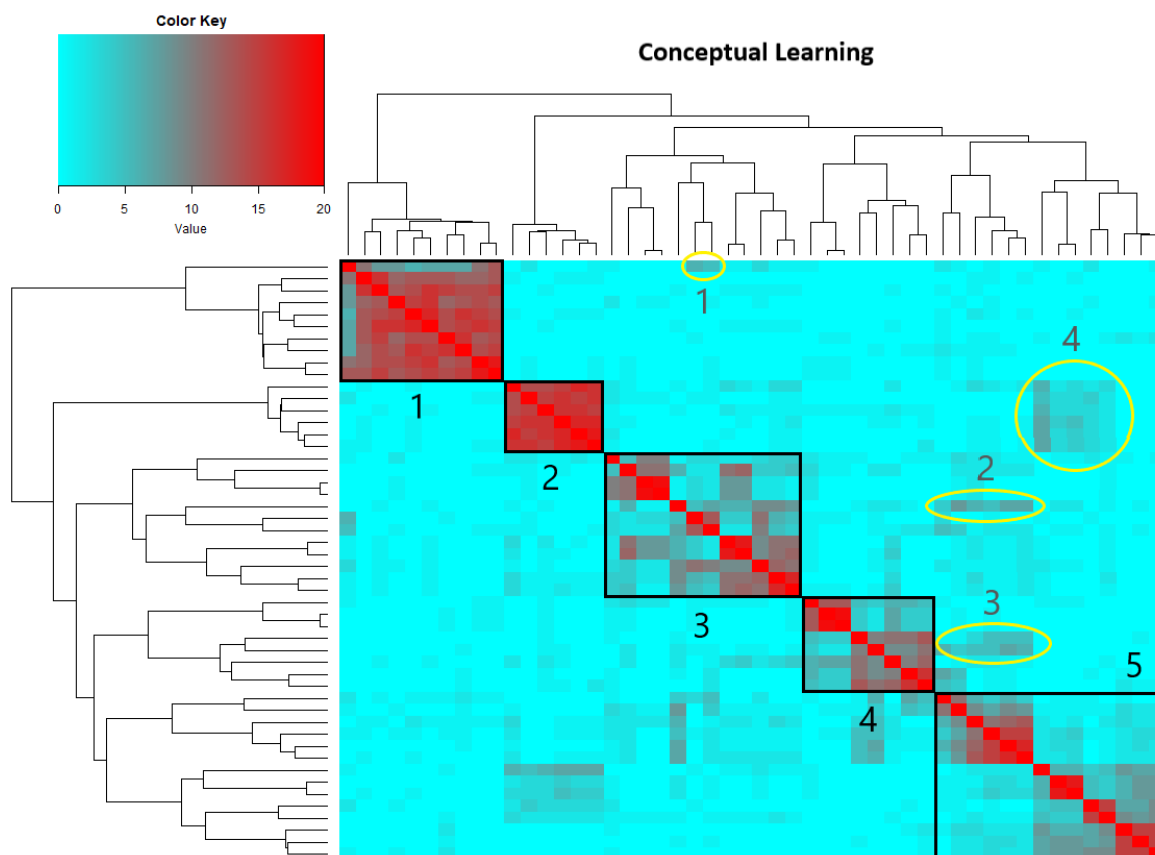


Figure 2 Heatmap showing word clusters at the diagonal based on card sorting ($N = 20$)

Table 1

Five clusters as taken from the heatmap

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Items	grip	weekend	meetings	widow	stolen
	flame	hours	room	sons	charges
	melting	days	college	mothers	expensive
	absorb	month	school	banker	shillings
	liquid	Friday	rented	owner	dollar
	fluid	evening	driveway	maid	cost
	mixture		parking	defendant	intervals
	solid		airport	sheriff	ten
	smooth		hotel		twenty
	soft		arrive		next
			trip		last
			vacation		lowest
					maximum
					plus

Table 2

Initial categories based on Huth et al. (2016)

Category	Tactile	Time	Place	Person	Number
Items	flame	days	hotel	sheriff	shillings
	soft	next	arrive	owner	intervals
	grip	weekend	driveway	maid	maximum
	fluid	hours	room	widow	plus
	absorb	trip	rented	banker	lowest
	melting	vacation	college	defendant	dollar
	smooth	Friday	school	sons	cost
	liquid	evening	airport	mothers	expensive
	mixture	month	parking	charges	ten
	solid	last	meetings	stolen	twenty

Comparing the found clusters (Table 1) with the categories based on Huth et al. (2016, Table 2), one can see that cluster 1 fully resembles the category “tactile” as the same items are included in both. Only the word “grip” showed ambiguity, which is indicated by the blue top line within the cluster and circle 1 in the heatmap (Figure 2). Circle 1 shows that some participants found that “grip” was also similar to the items “driveway” and “parking” from cluster 3.

Cluster 2 only partly overlaps with the category “time”. Missing are the items “next”, “last”, “trip” and “vacation”. Furthermore, this cluster shows a high similarity between all its items as no blue spots can be seen within the cluster.

The third cluster resembles the category “place” but additionally includes the items “trip” and “vacation” from category “time”. However, compared to the first two clusters, this one is scattered within. Similarity is not high between all items but rather between a few words. For example, (1) “meetings”, “room”, “college” and “school” seem to form their own subgroup which might be called “education”, so do (2) “airport”, “hotel”, “arrive”, “trip” and “vacation” as a subgroup “holiday”. The word “rented” shows similarity to items outside its cluster, namely with cluster 5 (Circle 2). This includes “stolen”, “charges”, “expensive”, “shillings”, “dollar” and “cost”.

Cluster 4 mostly overlaps with the category “person” except for the items “stolen” and “charges”. Furthermore, two subgroups can be seen made up of (1) “widow”, “sons” and “mothers” which seem to represent a subcategory “family” and (2) “banker”, “owner”, “maid”, “defendant” and “sheriff” which might represent a subcategory called “profession”. Additionally, circle 3 shows that “banker” and “owner” share similarities with “stolen”, “charges”, “expensive”, “shillings”, “dollar” and “cost” from cluster 5.

Lastly, the fifth cluster resembles category “number” but also includes the items “stolen”, “charges”, “next” and “last”. Again two subgroups can be made out in the heatmap: (1) “stolen”, “charges”, “expensive”, “shillings”, “dollar” and “cost” representing a subcategory “money” and (2) “intervals”, “ten”, “twenty”, “next”, “last”, “lowest”, “maximum” and “plus” as a subcategory “arithmetic”. Moreover, the second subgroup can even be divided further in a way that (2a) “intervals”, (2b) “ten” and “twenty”, (2c) “next” and “last”, and (2d) “lowest”, “maximum” and “plus” form their own groups within. Circle 4 shows that participants found that items “intervals”, “ten”, “twenty”, “next” and “last” also have similarities with cluster 2.

3.2 Questionnaire

The second part of the experiment concerned the questionnaire. This tested how well the categories fit the items assigned by Huth et al. (2016). Results present mean fit and standard deviation per category (Table 3). Further mean fit and standard deviation for each item is displayed in Tables 4-9.

Looking at the mean fit per category, all had a high score with above 3.00. Only category “tactile” had a score below 3.00 ($M = 2.47$). Which means that the items have a weaker but still a moderate relationship with this category. Low standard deviations show that the variance of the scores was small. So, ratings done by the participants did not differ much and were consistent.

In contrast to the regular categories, items used in the filler category showed the weakest relationship with the categories they were presented with ($M = 1.44$). Further, a paired-samples t-test was conducted to test if the filler items score significantly lower than the correct category items. The results confirmed that there was a significant difference between the scores; $t(19) = 11.606, p < .001$.

Table 3

Mean fit per category

	<i>N</i>	<i>M</i>	<i>SD</i>
Number	20	3.21	.46
Tactile	20	2.47	.64
Time	20	3.27	.42
Place	20	3.29	.34
Person	20	3.08	.48
Filler	20	1.44	.51

3.2.1 Mean ratings for category items

Table 4 displays mean ratings for the items of the first category “number”. Generally, there was a high fit between words and their category as their score was 2.95 or above. Especially, items “ten” ($M = 3.80$) and “twenty” ($M = 3.80$) stood out with scores close to 4.00. Only the item “shillings” had a moderate fit with $M = 2.60$. Regarding the standard deviations, half of the items (“shillings”, “maximum”, “plus”, “lowest”, “expensive”) had a high variance with above 0.80, so participants showed dissimilar ratings. The other half (“intervals”, “dollar”, “cost”, “ten”, “twenty”) had less variance in their ratings as standard

deviation were 0.64 and below. Again, words “ten” and “twenty” stood out with $SD = 0.41$ and $SD = 0.22$, respectively.

Table 4

Mean fit per item of category “number”

	<i>N</i>	<i>M</i>	<i>SD</i>
shillings	20	2.60	.94
intervals	20	3.10	.64
maximum	20	3.20	.89
plus	20	3.15	.88
lowest	20	3.30	.86
dollar	20	3.00	.65
cost	20	2.95	.60
expensive	20	3.00	.86
ten	20	3.80	.41
twenty	20	3.95	.22

Items identified as belonging to the category of “tactile” had mixed mean ratings (Table 5). The items “absorb” ($M = 1.45$) and “mixture” ($M = 1.5$) had the lowest fit with the category. A moderate fit was indicated for the words “flame” ($M = 2.05$), “fluid” ($M = 2.6$), “mixture” ($M = 2.25$) and “liquid” ($M = 2.55$). Remaining items had a high fit with a mean score of 3.00 or 3.10. Further, high standard deviations underline the mixed results as ratings between participants showed inconsistency.

Table 5

Mean fit per item of category “tactile”

	<i>N</i>	<i>M</i>	<i>SD</i>
flame	20	2.05	.94
soft	20	3.10	.79
grip	20	3.00	.79
fluid	20	2.60	.94
absorb	20	1.45	1.10
melting	20	2.25	.97
smooth	20	3.10	.91
liquid	20	2.55	1.05
mixture	20	1.50	1.05
solid	20	3.10	.91

For the third category “time”, the items “days”, “weekend”, “hours”, “Friday”, “evening”, and “month” were rated fitting well with mean scores above 3.50. Items “next”, “trip”, “vacation”, and “last” showed a lower rating below 3.00 which indicates a moderate fit. Ratings between participant were generally similar as indicated by low standard deviations. However, for the items “next” ($SD = 0.91$) and “last” ($SD = 1.12$) ratings showed more variance.

Table 6

<i>Mean fit per item of category “time”</i>			
	<i>N</i>	<i>M</i>	<i>SD</i>
days	20	3.60	.50
next	20	2.75	.91
weekend	20	3.50	.51
hours	20	3.75	.44
trip	20	2.80	.77
vacation	20	2.90	.79
friday	20	3.50	.69
evening	20	3.55	.60
month	20	3.55	.51
last	20	2.75	1.12

Table 7 presents scores for the items of category “place”. Overall, the words show a high fit with their category as their scores were above 3.00. Items “rented” ($M = 2.85$) and “meetings” ($M = 2.75$) had a somewhat lower score but were still in the upper mid-range. Low standard deviations show that participant’s ratings had small variance and were similar. Only the scores for word “rented” ($SD = 0.88$) showed higher inconsistency within the sample.

Table 7

Mean fit per item of category "place"

	<i>N</i>	<i>M</i>	<i>SD</i>
hotel	20	3.65	.49
arrive	20	3.35	.67
driveway	20	3.10	.72
room	20	3.60	.60
rented	20	2.85	.88
college	20	3.45	.60
school	20	3.50	.51
airport	20	3.45	.60
parking	20	3.20	.62
meetings	20	2.75	.79

Item-category fits for category "person" are presented below in Table 8. Again, overall ratings for items of this category were high with most scores being above 3.00. Exceptions to this are the words "charges" ($M = 2.15$) and "stolen" ($M = 1.45$) which stood out with a much lower category fit. Standard deviations within the category were mixed. Ratings for the items "sheriff", "owner", "maid", and "banker" were more consistent with standard deviations ranging from 0.50 to 0.60. The remaining six items had a higher variance in their rating as their standard deviation were above 0.80. Especially for the word "widow" ($SD = 1.24$) ratings were inconsistent.

Table 8

Mean fit per item of category "person"

	<i>N</i>	<i>M</i>	<i>SD</i>
sheriff	20	3.55	.60
owner	20	3.35	.59
maid	20	3.45	.60
widow	20	3.20	1.24
banker	20	3.55	.51
defendant	20	3.35	.81
sons	20	3.30	.86
mothers	20	3.40	.82
charges	20	2.15	.99
stolen	20	1.45	.89

3.2.2 Mean ratings for filler items

Lastly, Table 9 lists the mean category fit for the filler items. Almost all items had a low score with below or around 2.00. This illustrates a weak fit with the five categories used in the experiment they were assigned to. Item “husband” stands out with a high score of 3.60, meaning that it fits well with the assigned category. This is a logical result as the item was paired with the category “person” (see Appendix D, Question 55). Overall high standard deviations show that there was inconsistency in ratings between the participants.

Table 9

<i>Mean fit per filler item</i>			
	<i>N</i>	<i>M</i>	<i>SD</i>
plastic	20	.80	1.01
container	20	1.20	.89
steel	20	1.05	.69
rack	20	1.55	1.15
husband	20	3.60	.60
parents	20	1.50	1.19
family	20	2.20	1.11
cousin	20	.95	.76
hoping	20	.95	.10
paused	20	.75	.85
drifted	20	2.30	1.13
flooded	20	1.75	.97
disease	20	1.10	1.02
vile	20	2.15	1.27
cruelty	20	.90	.98
sorrow	20	1.85	1.14
mortals	20	1.30	1.03
eternal	20	1.05	.89
poisoning	20	.85	1.09
supposedly	20	1.00	1.03

4. Discussion

The present thesis aimed at finding out which way of categorization, based on prototypes or exemplars, applies best. For this, the perspective of brain activation was chosen as a basis. Huth et al. (2016) found semantic representations which presented eleven categories. The current experiment used 50 words from five categories out of the eleven to see if Huth et al.'s findings are able to predict manual categorization.

Results from the card sorting task showed that initial categories were only partly replicated. Only category "tactile" was reproduced completely by participants. Ambiguity of words also illustrated that category boundaries do not appear clear-cut but rather fuzzy. Furthermore, categorization behaviour showed that an exemplar-based way was preferred. This can be seen in the way that initial word groups were split up into subgroups and items were mixed up to fit smaller, more specific groups. For example, category "person" was split into subcategories "family" and "profession". It shows that participants did not categorize based on overall similarity but rather on what sets concepts apart. Within the category, items "banker" and "mothers" both represent persons however they differ in what they do as a person. A banker works with money while mothers care for a family. This difference was seen as more important than the similarity of the items, which further shows the use of specific exemplars while categorizing.

This result confirms Rouder and Ratcliff's (2006) speculation that for word categorization exemplars are used. In contrast to that, a prototype-based way would have yielded bigger word groups as there are degrees to membership (Aerts et al., 2016; Minda & Smith, 2001). So, atypical items would still belong to a bigger group and would not need a new subgroup made for them. Large groups like this were found by Huth et al. (2016) based on brain activation, however manual categorization did not yield the same results.

As mentioned above, category "tactile" was fully replicated while other categories were not. This could be because the category stands out from the others meaning that its items are less ambiguous and do not fit with outside items. Additionally, this could mean that this category was reproduced not because items belong together but because they did not fit with the others and were sorted out. Further, based on ratings from the questionnaire, this category had the lowest overall fit, so that the relationship with its items was only moderate. This shows that not all items fit the category even though they were grouped together by participants.

Comparing the card sorting task and the ratings from the questionnaire here, the results seem incompatible. Category "tactile" was reproduced completely but had the lowest

item-category fit. Because of that, another reason for that result might be that items have high similarity but that a different category (name) would have a better fit. In comparison, the word “tactile” does not appear as clear or straightforward in its meaning as for instance “person”.

Regarding the other categories, smaller groups were made by participants. As stated before, category “person” was broken up to create subcategories “family” and “profession”. Scores from the questionnaire show that the included items are highly related to their initial category “person”. Thus, when categorizing, participants preferred to make more specific groups but also validate the proposed category by Huth et al. (2016) as it can be considered a supergroup. An example for ambiguity of the items are the words “banker” and “owner”. Both were grouped into the subcategory “profession” but also showed similarity to the subcategory “money”, indicating fuzzy category boundaries. This seems logical as both occupations deal with money-related issues.

Also, category “number” was split up. Here, a subcategory “money” was made which additionally included the items “stolen” and “charges” from category “person”. Participants found that these items better belong to a specific concept (“money”) instead of atypical words belonging to category “person”. This result was also observed in the scores taken from the questionnaire where both items showed the lowest fit with the category “person”. Moreover, the second subcategory “arithmetic” made indicated that further division within this group is possible. In addition to that, the items “next” and “last” from category “time” were include here. Such as “stolen” and “charges”, these items were placed into a different better fitting group. An explanation for those changes could be that the current experiment gave no context with the to be categorized items while the narrative stories used in Huth et al. (2016) did. Here, a difference can be made between internal and external context (Galotti, 2004). In the current experiment internal context applies which relies on the subjective perspective of the participant. External context applies in Huth et al.’s study as the context was given by the environment; the narrative stories.

As observed for the previous category “person”, category “number” showed a high fit with its items as well. Again, participants validated Huth et al.’s (2016) category even though they created smaller groups in the card sorting task. This finding can be explained by work from Malt and Smith (1984) which showed that people detect within-category relationships. Participants were able to form small specific groups and still rate item-category fits for the initial categories high because the small groups are not new categories but belong to a more general one. Supplementary, this shows that prototypes, which are summary representations of a category, do not apply in this case as they do not include within-category differences.

Summarizing, the results give an answer to the research questions posed in this thesis. Regarding the first question it was found that semantic representations in the brain are overall not able to predict manual word categorization. However, item-category fit was generally rated high, so that categories proposed by Huth et al. (2016) remain relevant and thus should not be disregarded. The found semantic representations are able to give a general categorization structure but do not account for details within that structure. Furthermore, concerning the second research question, the results showed that participants created small and specific groups instead of larger general ones which indicates that word categorization is based on exemplars and not prototypes.

4.1 Limitations and recommendations for further study

As the current research used written words instead of narrative stories participants experienced the items differently. Context was removed which resulted in more ambiguity so that items were placed into different groups than the ones found by Huth et al. (2016). A future study including the contexts as given by the narrative stories might find altered results. Moreover, the sample of 50 words in the current experiment only covers a snapshot of the 10.470 words analysed by Huth et al. (2016). This presents a limitation of the chosen card sorting task as including such a high number of items can not be realised and would overwhelm participants. A machine learning based classifier would be able to work with this amount of data. Zubek and Kuncheva (2018) propose that psychology and machine learning are enriching each other and should continue doing so, especially in the area of categorization. Because of that, Huth et al.'s results should be applied or combined with methods from machine learning. As the current experiment found a high fit between items and their categories, this application could give more insights.

Furthermore, the finding that participants use exemplars and create smaller, more specific groups can help when designing websites. For example, browsing and searching by users can be supported (Feldman, 2004). Organizing content into smaller groups gives a better overview and aids search speed. Groups should be created based on high typicality of the items to make it more intuitive for users. Additionally, labels for the groups should be clear and unambiguous. This is supported by the finding of the current research in which the category (name) "tactile" was rated less fitting with its items even though the group was replicated completely.

4.2 Conclusion

Concluding, the goal of this thesis was to find out if semantic representations in the brain could predict word categorization and show if exemplars or prototypes are used. Results showed that semantic representations were only partly able to predict categorization. Participants preferred smaller specific groups than the ones based on brain activation. Because of this, exemplar-based categorization applies when grouping words.

5. References

- Aerts, D., Broekaert, J., Gabora, L., & Sozzo, S. (2016). Generalizing prototype theory: A formal quantum framework. *Frontiers In Psychology*, 7. doi:10.3389/fpsyg.2016.00418
- Barsalou, L. (1989). Intraconcept similarity and its implications for interconcept similarity. *Similarity And Analogical Reasoning*, 76-121. doi:10.1017/cbo9780511529863.006
- Bellezza, F. (1984). Reliability of retrieval from semantic memory: Noun meanings. *Bulletin Of The Psychonomic Society*, 22(5), 377-380. doi:10.3758/bf03333850
- Cai, Y., Au Yeung, C., & Leung, H. (2012). Concepts and categorization from a psychological perspective. *Fuzzy Computational Ontologies In Contexts*, 23-35. doi:10.1007/978-3-642-25456-7_3
- Cooke, N. (1994). Varieties of knowledge elicitation techniques. *International Journal Of Human-Computer Studies*, 41(6), 801-849. doi: 10.1006/ijhc.1994.1083
- Cruse, D. (2001). Lexical semantics. *International Encyclopedia Of The Social & Behavioral Sciences*, 8758-8764. doi:10.1016/b0-08-043076-7/02990-9
- Dopkins, S., & Gleason, T. (1997). Comparing exemplar and prototype models of categorization. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 51(3), 212–230. doi:10.1037/1196-1961.51.3.212
- Feldman, S. (2004). *Why categorize?* Retrieved from: <https://www.kmworld.com/Articles/Editorial/Features/Why-categorize-9580.aspx>
- Gallantlab.org (n.d.) *Semantic map*. Retrieved from: <http://gallantlab.org/huth2016/>
- Galotti, K. M. (2004). *Cognitive psychology in and out of the laboratory* (3rd ed). Belmont, CA, USA: Thomson/Wadsworth
- Grossman, M., Smith, E., Koenig, P., Glosser, G., DeVita, C., Moore, P., & McMillan, C. (2002). The neural basis for categorization in semantic memory. *Neuroimage*, 17(3), 1549-1561. doi:10.1006/nimg.2002.1273
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453-458. doi:10.1038/nature17637
- Huth, A., Nishimoto, S., Vu, A., & Gallant, J. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, 76(6), 1210-1224. doi:10.1016/j.neuron.2012.10.014
- Jones, M., Willits, J., & Dennis, S. (2015). Models of semantic memory. *Oxford Handbooks Online*. doi:10.1093/oxfordhb/9780199957996.013.11

- Malt, B., & Smith, E. (1984). Correlated properties in natural categories. *Journal Of Verbal Learning And Verbal Behavior*, 23(2), 250-269. doi:10.1016/s0022-5371(84)90170-1
- McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, 6(4), 462-472. doi:10.3758/bf03197480
- Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, 35, 113–138. doi:10.1146/annurev.ps.35.020184.000553
- Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3), 775–799. doi:10.1037/0278-7393.27.3.775
- Murphy, G. (2016). Is there an exemplar theory of concepts? *Psychonomic Bulletin & Review*, 23(4), 1035-1042. doi:10.3758/s13423-015-0834-3
- Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 18–39). Cambridge, UK: Cambridge University Press.
- Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(5), 924-940. doi:10.1037/0278-7393.28.5.924
- Patterson, K., Nestor, P., & Rogers, T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976-987. doi:10.1038/nrn2277
- Rosch, E. (1983). Prototype classification and logical classification: The two systems. In E. F. Scholnick (Ed.), *New trends in conceptual representation: Challenges to Piaget's theory?* (pp. 73–86). New York, NY, USA: Erlbaum.
- Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382-439. doi:10.1016/0010-0285(76)90013-x
- Rouder, J. N., & Ratcliff, R. (2006). Comparing exemplar and rule-based theories of categorization. *Current Directions in Psychological Science*, 15, 9–13. doi:10.1111/j.0963-7214.2006.00397.x.
- Schmettow, M., & Sommer, J. (2016). Linking card sorting to browsing performance – are congruent municipal websites more efficient to use? *Behaviour & information technology*, 35(6), 452-470. doi:10.1080/0144929X.2016.1157207

- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA, USA: Harvard University Press.
- Torre, G. (2017). The brain's building blocks: Of protons and voxels. Retrieved from: <https://knowingneurons.com/2017/09/27/mri-voxels/#:~:text=Re%2Denter%20the%20voxel%3A%20A,created%20by%20protons%20magnet%20interactions.>
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson, *Organization of memory* (pp. 381-403). Cambridge, MA, USA: Academic Press.
- Typicality effect. (2020). In *APA Dictionary of Psychology*. Retrieved from <https://dictionary.apa.org/typicality-effect>
- Voorspoels, W., Vanpaemel, W., & Storms, G. (2008). Exemplars and prototypes in natural language concepts: A typicality-based evaluation. *Psychonomic Bulletin & Review*, 15(3), 630-637. doi:10.3758/pbr.15.3.630
- Zubek, J., & Kuncheva, L. (2018). *Learning from exemplars and prototypes in machine learning and psychology*. Retrieved from: <https://arxiv.org/abs/1806.01130>

6. Appendices

Appendix A

Informed Consent Form

Informed Consent

Dear participant,

This study aims to gain information about how concepts and conceptual spaces are learned. Therefore, it involves a card sorting task and completing a questionnaire afterwards. This will take approximately 40 minutes.

If you have any questions or concerns about this study, you can contact me at p.l.elsasser@student.utwente.nl

All data is kept anonymously and personal information will not be passed on to third parties under any condition. Under no circumstances will any personal data or identifying information be included in the report of this research. Nobody, except the researcher and the supervisor will have access to the anonymized data in its entirety. Participation in this study is voluntarily and you can withdraw at any time. This research project has been reviewed and approved by the BMS Ethics Committee.

By signing this, you declare the following:

I have read and understood the study information dated and I have been able to ask questions about the study. Further, I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason. I understand that taking part in the study involves a card sorting task and filling in a questionnaire, and that information I provide will be used for study and research purposes only. Additionally, I understand that personal information about me (e.g. gender, age) will not be shared.

.....
Location, Date

.....
Signature participant

Appendix B

Word list (card sorting) and filler items based on findings of Huth et al. (2016)

Word List

Word Number	Word	Category	Voxel Number	Location	Reliability Score
1	shillings	Number	[18,30,25]	RH, frontal lobe	Good, very reliable
2	intervals	Number	[21,70,27]	RH, parietal lobe	Good, very reliable
3	maximum	Number	[21,70,27]	RH, parietal lobe	Good, very reliable
4	plus	Number	[21,70,27]	RH, parietal lobe	Good, very reliable
5	lowest	Number	[21,70,27]	RH, parietal lobe	Good, very reliable
6	dollar	Number	[16,25,72]	LH, frontal lobe	Good, very reliable
7	cost	Number	[16,25,72]	LH, frontal lobe	Good, very reliable
8	expensive	Number	[16,25,72]	LH, frontal lobe	Good, very reliable
9	ten	Number	[14,89,63]	LH, occipital lobe	Excellent, extremely reliable
10	twenty	Number	[14,89,62]	LH, occipital lobe	Excellent, extremely reliable
11	flame	Tactile	[18,69,77]	LH, parietal lobe	Excellent, extremely reliable
12	soft	Tactile	[19,66,78]	LH, parietal lobe	Excellent, extremely reliable
13	grip	Tactile	[19,66,78]	LH, parietal lobe	Excellent, extremely reliable
14	fluid	Tactile	[18,69,77]	LH, parietal lobe	Excellent, extremely reliable
15	absorb	Tactile	[18,69,77]	LH, parietal lobe	Excellent, extremely reliable
16	melting	Tactile	[21,67,25]	RH, parietal lobe	Good, very reliable
17	smooth	Tactile	[21,67,25]	RH, parietal lobe	Good, very reliable
18	liquid	Tactile	[21,67,25]	RH, parietal lobe	Good, very reliable
19	mixture	Tactile	[21,67,25]	RH, parietal lobe	Good, very reliable
20	solid	Tactile	[21,67,25]	RH, parietal lobe	Good, very reliable

21	days	Time	[6,41,23]	RH, temporal lobe	Good, very reliable
22	next	Time	[6,41,23]	RH, temporal lobe	Good, very reliable
23	weekend	Time	[6,41,23]	RH, temporal lobe	Good, very reliable
24	hours	Time	[6,41,23]	RH, temporal lobe	Good, very reliable
25	trip	Time	[6,41,23]	RH, temporal lobe	Good, very reliable
26	vacation	Time	[15,76,26]	RH, occipital lobe	Good, very reliable
27	friday	Time	[15,89,62]	LH, occipital lobe	Good, very reliable
28	evening	Time	[18,15,42]	RH, frontal lobe	Good, very reliable
29	month	Time	[22,80,37]	RH, parietal lobe	Good, very reliable
30	last	Time	[22,80,37]	RH, parietal lobe	Good, very reliable
31	hotel	Place	[24,32,35]	RH, frontal lobe	Good, very reliable
32	arrive	Place	[24,32,35]	RH, frontal lobe	Good, very reliable
33	driveway	Place	[24,32,35]	RH, frontal lobe	Good, very reliable
34	room	Place	[17,72,24]	RH, parietal lobe	Excellent, extremely reliable
35	rented	Place	[14,88,64]	LH, occipital lobe	Excellent, extremely reliable
36	college	Place	[14,88,64]	LH, occipital lobe	Excellent, extremely reliable
37	school	Place	[14,88,64]	LH, occipital lobe	Excellent, extremely reliable
38	airport	Place	[15,89,61]	LH, occipital lobe	Good, very reliable
39	parking	Place	[15,89,61]	LH, occipital lobe	Good, very reliable
40	meetings	Place	[18,15,42]	RH, frontal lobe	Good, very reliable
41	sheriff	Person	[15,81,29]	RH, occipital lobe	Excellent, extremely reliable
42	owner	Person	[15,81,29]	RH, occipital lobe	Excellent, extremely reliable
43	maid	Person	[15,81,29]	RH, occipital lobe	Excellent, extremely reliable
44	widow	Person	[24,27,40]	RH, frontal lobe	Good, very reliable

45	banker	Person	[24,27,40]	RH, frontal lobe	Good, very reliable
46	defendant	Person	[16,81,68]	LH, parietal lobe	Good, very reliable
47	sons	Person	[15,86,68]	LH, parietal lobe	Good, very reliable
48	mothers	Person	[16,81,68]	LH, parietal lobe	Good, very reliable
49	charges	Person	[22,39,58]	LH, parietal lobe	Excellent, extremely reliable
50	stolen	Person	[22,39,58]	LH, parietal lobe	Excellent, extremely reliable

Filler List

Word Number	Word	Category	Voxel Number	Location	Reliability Score
51	plastic	Visual	[19,37,73]	LH, frontal lobe	Excellent, extremely reliable
52	container	Visual	[19,37,73]	LH, frontal lobe	Excellent, extremely reliable
53	steel	Visual	[19,37,73]	LH, frontal lobe	Excellent, extremely reliable
54	rack	Visual	[19,37,73]	LH, frontal lobe	Excellent, extremely reliable
55	husband	Social	[15,79,27]	RH, occipital lobe	Good, very reliable
56	parents	Social	[15,79,27]	RH, occipital lobe	Good, very reliable
57	family	Social	[17,74,25]	RH, parietal lobe	Excellent, extremely reliable
58	cousin	Social	[17,74,25]	RH, parietal lobe	Excellent, extremely reliable
59	hoping	Mental	[21,20,38]	RH, frontal lobe	Good, very reliable
60	paused	Mental	[21,20,38]	RH, frontal lobe	Good, very reliable
61	drifted	Outdoor	[16,15,32]	RH, frontal lobe	Excellent, extremely reliable
62	flooded	Outdoor	[16,15,32]	RH, frontal lobe	Excellent, extremely reliable
63	disease	Violence	[12,55,74]	LH, temporal lobe	Excellent, extremely reliable
64	vile	Violence	[14,33,74]	LH, frontal lobe	Good, very reliable

65	cruelty	Violence	[14,33,74]	LH, frontal lobe	Good, very reliable
66	sorrow	Mental	[11,52,22]	RH, temporal lobe	Excellent, extremely reliable
67	mortals	Mental	[11,52,22]	RH, temporal lobe	Excellent, extremely reliable
68	eternal	Mental	[11,52,22]	RH, temporal lobe	Excellent, extremely reliable
69	poisoning	Violence	[10,54,74]	LH, temporal lobe	Excellent, extremely reliable
70	supposedly	Violence	[10,54,74]	LH, temporal lobe	Excellent, extremely reliable

Appendix C

Questionnaire

The questionnaire now asks you to rank word pairs on how well they relate. You can choose on a range from "highly related" to "highly not related".

Q1 How do you judge the relation of this pair of words: shillings - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q2 How do you judge the relation of this pair of words: intervals - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q3 How do you judge the relation of this pair of words: maximum - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q4 How do you judge the relation of this pair of words: plus - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q5 How do you judge the relation of this pair of words: lowest - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q6 How do you judge the relation of this pair of words: dollar - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q7 How do you judge the relation of this pair of words: cost - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q8 How do you judge the relation of this pair of words: expensive - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q9 How do you judge the relation of this pair of words: ten - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q10 How do you judge the relation of this pair of words: twenty - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q11 How do you judge the relation of this pair of words: flame - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q12 How do you judge the relation of this pair of words: soft - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q13 How do you judge the relation of this pair of words: grip - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q14 How do you judge the relation of this pair of words: fluid - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q15 How do you judge the relation of this pair of words: absorb - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q16 How do you judge the relation of this pair of words: melting - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q17 How do you judge the relation of this pair of words: smooth - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q18 How do you judge the relation of this pair of words: liquid - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q19 How do you judge the relation of this pair of words: mixture - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q20 How do you judge the relation of this pair of words: solid - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q21 How do you judge the relation of this pair of words: days - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q22 How do you judge the relation of this pair of words: next - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q23 How do you judge the relation of this pair of words: weekend - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q24 How do you judge the relation of this pair of words: hours - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q25 How do you judge the relation of this pair of words: trip - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q26 How do you judge the relation of this pair of words: vacation - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q27 How do you judge the relation of this pair of words: friday - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q28 How do you judge the relation of this pair of words: evening - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q29 How do you judge the relation of this pair of words: month - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q30 How do you judge the relation of this pair of words: last - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q31 How do you judge the relation of this pair of words: hotel - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q32 How do you judge the relation of this pair of words: arrive - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q33 How do you judge the relation of this pair of words: driveway - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q34 How do you judge the relation of this pair of words: room - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q35 How do you judge the relation of this pair of words: rented - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q36 How do you judge the relation of this pair of words: college - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q37 How do you judge the relation of this pair of words: school - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q38 How do you judge the relation of this pair of words: airport - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q39 How do you judge the relation of this pair of words: parking - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q40 How do you judge the relation of this pair of words: meetings - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q41 How do you judge the relation of this pair of words: sheriff - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q42 How do you judge the relation of this pair of words: owner - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q43 How do you judge the relation of this pair of words: maid - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q44 How do you judge the relation of this pair of words: widow - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q45 How do you judge the relation of this pair of words: banker - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q46 How do you judge the relation of this pair of words: defendant - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q47 How do you judge the relation of this pair of words: sons - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q48 How do you judge the relation of this pair of words: mothers - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q49 How do you judge the relation of this pair of words: charges - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q50 How do you judge the relation of this pair of words: stolen - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q51 How do you judge the relation of this pair of words: plastic - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q52 How do you judge the relation of this pair of words: container - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q53 How do you judge the relation of this pair of words: steel - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q54 How do you judge the relation of this pair of words: rack - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q55 How do you judge the relation of this pair of words: husband - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q56 How do you judge the relation of this pair of words: parents - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q57 How do you judge the relation of this pair of words: family - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q58 How do you judge the relation of this pair of words: cousin - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q59 How do you judge the relation of this pair of words: hoping - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q60 How do you judge the relation of this pair of words: paused - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q61 How do you judge the relation of this pair of words: drifted - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q62 How do you judge the relation of this pair of words: flooded - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q63 How do you judge the relation of this pair of words: disease - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q64 How do you judge the relation of this pair of words: vile - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q65 How do you judge the relation of this pair of words: cruelty - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q66 How do you judge the relation of this pair of words: sorrow - time

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q67 How do you judge the relation of this pair of words: mortals - place

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q68 How do you judge the relation of this pair of words: eternal - tactile

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q69 How do you judge the relation of this pair of words: poisoning - number

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Q70 How do you judge the relation of this pair of words: supposedly - person

- ☐ highly related
 - ☐ related
 - ☐ neutral
 - ☐ not related
 - ☐ highly not related
-

Appendix D

R script for vector analysis of clusters

```
# R script to generate a heatmap
```

```
# Call these libraries. They need to be installed as packages
```

```
library(gplots)
```

```
library(RColorBrewer)
```

```
# Read the data file
```

```
data <- read.csv("path/filename")
```

```
# Transform data in numerical format
```

```
mat_data <- data.matrix(data[,1:ncol(data)])
```

```
# Define colors of heatmap: red for high numbers
```

```
my_palette <- colorRampPalette(c("cyan", "red"))(n = 299)
```

```
# Call heatmap function (from gplots), with these arguments
```

```
# See:
```

```
https://www.rdocumentation.org/packages/gplots/versions/3.0.1/topics/heatmap.2
```

```
# Note: argument 'main=' gives name of plot
```

```
heatmap.2(mat_data, col = my_palette, density.info="none", trace="none", revC = TRUE,  
main="Name")
```