UNIVERSITY OF BELGRADE

FACULTY OF PHILOSOPHY

Bojan R. Lalić

THE ROLE OF LABEL FEATURES AND LABEL REMEMBERING IN CONCEPT FORMATION: BEHAVIOURAL, NEURAL AND COGNITIVE MODELLING APPROACH

Doctoral Dissertation

Belgrade, 2019

УНИВЕРЗИТЕТ У БЕОГРАДУ

ФИЛОЗОФСКИ ФАКУЛТЕТ

Бојан Р. Лалић

УЛОГА СВОЈСТАВА И ПАМЋЕЊА ИМЕНИТЕЉА НА ФОРМИРАЊЕ ПОЈМОВА: БИХЕЈВИОРАЛНИ, НЕУРАЛНИ И ПРИСТУП КОГНИТИВНОГ МОДЕЛОВАЊА

докторска дисертација

Београд, 2019

Supervisor:
Associate professor, **Vanja Ković**, PhD,
University of Belgrade, Faculty of Philosophy

Board members:
Professor **Dejan Lalović**, PhD,
University of Belgrade, Faculty of Philosophy

Assistant professor **Dejan Pajić** PhD,
University of Novi Sad, Faculty of Philosophy

Assistant professor, **Predrag Tadić**, PhD,
University of Belgrade, School of electrical engineering

Professor **Gert Westermann**, PhD,
University of Lancaster, UK

Defence date: _______________

ACKNOWLEDGEMENTS

As a young student, I always thought that scientific work was a product of a single man. I thought that sitting and working by oneself was enough to deliver scientific achievements. Nothing could be further from the truth! Even a small step in scientific progress is the result of complementary work and cooperation with fellow researchers, however, it also depends on moral support from colleagues, friends and family.

Firstly, I would like to state my highest gratitude to my supervisor prof. dr Vanja Ković, who showed trust in my work, even when it was widely disputed. She also inspired me at some critical moments to continue my research, when I encountered troubles. Furthermore, she introduced me to new methodological approaches in psychology and to some very competent people in the area, which was priceless for my further personal and academic development.

Further, I owe great thanks and appreciation to the board members: the President of the board prof. dr Dejan Lalović, doc. dr Dejan Pajić, doc. dr Predrag Tadić, while a special note to prof. dr Gert Westermann, who accepted this position despite not being a resident of our country.

I owe great recognition to the contributors of scientific work conducted in this dissertation: prof. dr Suzy J Styles, without which linguistic competences, design of the verbal stimuli would not have been possible. I must further acknowledge dr Andrej Savić, whose technical knowledge and contribution to ERP recording and analysis was enormous. Finally, I owe gratitude to a PhD student Marija Novičić MSc (graduate engineer), who assisted with connectionist modelling simulations.

I also owe great thanks to my good friend Niall Ardill, who helped me with proofreading this dissertation.

I am also very grateful to the members of the Laboratory for Neurocognition and Applied Cognition at the Faculty of Philosophy in Belgrade, primarily to: Anđela, Jelena, Jadranka, Marija, Mina and Vojislav, who were always there to assist me with my experiments and to help me with all other work, especially during my absence.

Many thanks to doc. dr Lazar Tenjović and doc. dr Alesksandar Zorić who supported me in recruiting the last group of participants for my experiments from among their freshmen students.

Finally, I owe a great debt of gratitude to my dear participants (298 of them in total). Every experiment lasted for around one hour, which was sometimes exhausting for them, but their enthusiasm made my work much easier. The first result of this research was that almost all students within the psychology department knew which were ‘good’ and which were ‘bad’ aliens ☺.

Last but not least, I owe the greatest thanks to my dearest Snežana, who was always there to support me in some very difficult moments. Her unconditional support, patience and care were priceless. Additionally, she lent her beautiful voice for verbal stimuli in my experiments.

THE ROLE OF LABEL FEATURES AND LABEL REMEMBERING IN CONCEPT FORMATION: BEHAVIUORAL, NEURAL AND COGNITIVE MODELLING APPROACH

Abstract

The basic aim of this dissertation was to demonstrate the way in which different label features influence concept formation, specifically their learning and generalization. These labels could be verbal or non-verbal, presented auditory or visually. In the experiments, participants learned novel categories (aliens) labelled with novel labels (pseudo-words or novel non-verbal emblems and sounds). During experiments, ERPs were recorded.

In Chapter I, the effects of verbal auditory labels on concept formation were examined, specifically, the way in which the level of phonological difference between labels influence the learning of novel categories. Results showed that learning of categories labelled with phonologically more different labels was significantly faster and generalized better compared to categories labelled with phonologically more similar labels. Furthermore, this property is independent from the effects of sound symbolism.

In Chapter II, the effects of differences of non-verbal labels (visual or auditory) on category learning was examined. Results showed that there were no differences of influence of non-verbal labels on category learning, no matter if they were more different or not.

In Chapter III, relations between the effects of verbal and non-verbal labels on category learning were examined. Results showed that auditory verbal labels which were phonologically more different led to faster learning and generalization of novel categories, which was not the case with other types of labels.

In Chapter IV, the effects of explicit instruction given to the participants in the experiments to pay special attention to the labels during categorisation and to learn them were examined. Results showed that participants learned faster and generalized better once the instruction was given, while the absence of instruction led to the diminishing of label effects on category learning.

Finally, in Chapter V, a neural network was constructed, the task of which was to simulate the effects of phonological differences of labels on category learning. The model successfully simulated these differences, since it learned categories labelled with phonologically more different labels faster compared to the less different ones.

From the results obtained in this dissertation we can conclude that the effects of label features on category learning is significant, which is specially the case with auditory verbal labels and their phonological difference. As a result of these findings, category learning based on the difference level hypothesis was designed. This hypothesis explains category learning as probability, which is based on the compound difference between visual properties of the categories and the difference of their labels. These results lead to the more fundamental conclusion that relations between language and thought are mutually influential and that these entities are not completely independent.

Keywords: Categorisation, language and thought, category learning, category labels, label phonological differences, category learning based on the difference level, sound symbolism, evoked potentials, cognitive modelling.

Scientific field: Psychology
Scientific subfield: General and Cognitive psychology
UDC number: 159.953 (043.3)

УЛОГА СВОЈСТАВА И ПАМЋЕЊА ИМЕНИТЕЉА НА ФОРМИРАЊЕ ПОЈМОВА: БИХЕЈВИОРАЛНИ, НЕУРАЛНИ И ПРИСТУП КОГНИТИВНОГ МОДЕЛОВАЊА

Сажетак

Основни циљ ове дисертације је био да покаже како различите карактеристике именитеља утичу на формирање појмова, тј. на њихово учење и генерализацију. Ти именитељи могу бити вербални и невербални, изложени аудитивно или визуелно. Коришћени су експерименти у којима су испитаници учили нове категорије (ванземаљце) именоване новим именитељима (пресудо-речима или претходно невиђеним невербалним амблемима и звуковима), при чему су испитаницима снимани и ERP мождани таласи.

У I поглављу, испитивана је улога аудитивних вербалних именитеља на формирање појмова, тачније како фонолошка разлика именитеља утиче на учење нових категорија. Резултати су показали да је учење категорија именовано фонолошки различитијим именитељима брже, а генерализација боља у односу на оне именоване фонолошки сличнијим именитељима. Такође, ова особина је независна од утицаја језичког симболизма.

У II поглављу, испитиван је утицај различитости невербалних именитеља на учење категорија (визуелних и аудитивних). Резултати су показали да не постоје разлике у утицају невербалних именитеља на учење категорија, без обзира на то да ли су они различитији или не.

У III поглављу, испитиван је однос утицаја вербалних и невербалних именитеља на учење категорија. Резултати су показали да само аудитивни вербални именитељи који су фонолошки различити доводе до бржег учења и генералицације, што није случај са осталим врстама именитеља.

У IV поглављу је испитиван утицај експлицитне инструкције испитаницима да обрате пажњу на именитеље и да их запамте. Резултати су показали да испитаници брже уче и боље генерализују категорије када добију инструкцију, док изостанком инструкције ефекти на учење категорија изостају.

Најзад, у V поглављу је конструисана неурална мрежа чији је задатак био да симулира ефекте фонолошке различитости именитеља. Модел је успешно симулирао ове различитости, јер је показао да је учење брже када су фонолошке разлике између именитеља веће.

Резултати добијени у дисертацији наводе на закључак да је утицај карактеристика именитеља на усвајање појмова изузетно битан и да се пре свега односи на вербалне именитеље и њихове фонолошке разлике. Као производ ових резултата, дефинисана је хипотеза учења категорија базираног на нивоу разлике. Ова хипотеза описује учење категорија као вероватноћу која је базирана на заједничкој разлици између њихових визуелних карактеристика и разлика у именитељима. Такви резултати наводе на фундаменталнији закључак да је однос мишљења и језика међузависан и да они никако нису потпуно независни.

Кључне речи: Категоризација, мишљење и језик, учење категорија, именитељи категорија, фонолошке разлике именитеља, учење категорија базираног на нивоу разлике, језички симболизам, евоцирани потенцијали, когнитивно моделовање.

Научна област: Психологија
Ужа научна област: Општа и когнитивна психологија
УДК број: 159.953 (043.3)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I dedicate this dissertation to my late mother Veselinka and late grandmother Milica. The first brought me to this world and the second brought me up in it.

THE ROLE OF LABEL FEATURES AND LABEL REMEMBERING IN CONCEPT FORMATION: BEHAVIOURAL, NEURAL AND COGNITIVE MODELLING APPROACH

INTRODUCTION

1. CONCEPTS AND CATEGORIES

The world around us is enormously complex. Confirmation to this claim is the enormous number of objects that we are surrounded with. On the other hand, human mental capacity is limited, especially its working memory. With such limited resources it would be completely impossible to deal with such a complex world if we did not group similar objects into broader classes – **categories**. In the process of categorisation, similarities between objects are put forward, while their differences are neglected.

Mental representation of the category is a **concept** (Eysenck & Keane, 2000; Cohen & Lefebvre, 2005). Once category is learned, it forms a concept. Based on concepts, we can structure the world around us despite its enormous complexity. Concepts enable our working memory to cover more objects than its capacity could normally allow.

Theories about categories are not unified. Generally, we can divide these theories in two broad groups: *a classical view* and *a modern view* (Murphy, 2002; Cohen & Lefebvre, 2005). Additionally, we could add to this list a *connectionist view*, which belongs to the modern view, but sometimes is separately reported.

Classical view

Historically, the most dominant and enduring theory of categorisation was the theory of object attributes, which was rooted in Aristotelian theory of categorisation (Aristotle, 350BC/1965) and further specified by German philosopher Frege (Frege, 1879/1952). The main characteristics of this view are the following (Smith & Medin, 1981; Eysenck & Keane, 2000):

1. Categories are represented as a list of their attributes. Some of these attributes are necessary and some are sufficient. 2. Objects either belong to the category or not. There is no space between categories. 3. There are no differences within category in the sense of typicality. All members of the category are equally representative. 4. Categories are hierarchically organized. Lower members share all attributes with higher members.

From the broader point, this view of categories and its hierarchical organization is presented in a once very popular model: Quillian propositional approach (Quillian, 1968; Collins and Quillian, 1969).

The first who challenged this view was Ludwig Wittgenstein in his latter work “Philosophical investigations” (Wittgenstein, 1953). Firstly, Wittgenstein stressed that the categories are not always easily classifiable into one or another category. Additionally, it is not easy to give a complete list of attributes to some categories, since they are so diverse, but still easily identified as a member of a category (example he gives is category of games).

Modern view

The classical view on categories was substituted with new **prototypical view** which was developed by Eleanor Rosh (Rosch & Mervis, 1975; Rosch 1975; Rosch, 1978; Barsalou, 1985). This view considers that members of the category are not equal, but rather that there are members which are more typical and which are called prototypes.

What makes a category member more prototypical? That is not frequency of occurrence (since there are many frequent members which are not prototypical, like chicken for birds), but rather family resemblance. Family resemblance is calculated by the number of attributes that one member shares with all other members of the category. Additionally, typicality is determined by the number of attributes that a category member shares with other categories: if this number is lower, a category member is more typical. Merging of these two criteria is a good predictor of typicality which Rosh and Mervis empirically tested (Rosch & Mervis, 1975).

Exemplar based view. This view was developed by Medin and Schaffer (Medin & Schaffer, 1978) and it is based on the claim that there is no abstraction at all. The concept is representation of all previously encountered exemplars of the category. For example, the concept of dogs represents the collection of all previously encountered dogs we remember. This collection is based on token encounters (which are real exemplars), rather than type encounters. Likewise, the dog one meets in the neighbourhood every day will have the highest contribution to his representation, since each encounter represents an individual exemplar.

An important aspect of this theory is the computational model of similarity calculation which is based on the multiplicative rule (while previous Rosch's model was based on summation rule). Attributes matched between two objects which are represented with values 0 to 1 (0 – completely different, 1 – completely the same). These values are multiplied and overall similarity is calculated. As similarity is higher, two objects are categorized as the members of the same category. Likewise, it can happen that two objects are very similar, but they have one attribute very different (close to 0) and the overall similarity will be very small.

Finally, the last approach is called the **knowledge approach** and is rather an approach than a completed hypothesis (Murphy & Medin, 1985, Murphy, 2002). This view signifies influence of previous knowledge, which is not independent from the conceptual knowledge. So, changes in the view of one category can make changes in all other categories. For example, realizing that some snails are hermaphrodites can change all views of biological propagation.

This view is closely related with concept of schema (Bartlett, 1932), which represents an organized cluster of concepts which are mutually interdependent and placed in so called slots. These slots could be used or not. Change in any part of the schema, necessarily leads to changes in other parts of the schema, which represents interdependence.

Connectionist view

The connectionist view of parallel distributed processing – PDP (Rumelhart, McClelland & The PDP Research Group, 1986; McClelland & Rogers, 2003; McLeod, Plunkett & Rolls, 1998) considers concepts coded as connection weight between units which are organized in several layers (usually three). Layers are organized in input, hidden and output layer, and all are interconnected, except input and output layer which are indirectly connected over the hidden layer. Concept knowledge is not located in one place, but rather spread over several connections. Concept is coded in specific units' activation at the output.

These models managed to demonstrate categorization of objects very successfully. Apart from the previous view, this view is biologically plausible (if we consider units as individual neurons). Furthermore, it is resistant to damage (excluding several units or connections, still leads to correct classification) and successfully demonstrates the effects of some pathological processes. Finally, these models successfully managed to demonstrate categorical hierarchy, prototype effects and most of the experimental findings. For this reason, these models are very good in the cross-examination of experimental results, testing and generation of new hypothesis.

The shortcoming of these models is that there is no complete biological correspondence between the models and brain functions (for example, nothing like the most frequently used learning algorithm for these models – back propagation, was identified in brain functions).

2. CATEGORY LEARNING

Concepts are not innate. Internal representations of categories are created in the learning process. There are numerous types and classifications of learning, but here will be presented the most relevant theories of category learning.

Previously stated theories of categorization could also be assumed as learning theories. While for the prototype view category learning is conducted through the "building" of prototype representation, for an exemplar based view, the process of category learning is the process of learning individual exemplars.

All these theories view category learning as a single cognitive process. Novel research showed that learning is not a single process, but rather consists of several independent processes (Ashby & Valentin, 2017; Wahlheim, McDaniel & Little, 2016). One of the most developed theories in this group is the COVIS theory of category learning, proposed by Ashby (Ashby, Alfonso-Reese, Turken & Waldron, 1998; Ashby & Maddox 2005; Ashby & Maddox 2011).

COVIS (Competition between Verbal and Implicit Systems) considers the existence of two independent learning systems: the **Frontal System**, which is in charge for declarative learning and the **Basal Ganglia Mediated System**, for procedural learning. In examining these systems, two different kinds of tasks are usually used: Rule-Based (RB) and Information Integration (II). The RB task is based on a single (or several) explicit rule(s) which are clear-cut between categories and which can be easily verbally described. The II tasks are based on the difference of multiple incommensurable dimensions, which are not easy for verbal description.

Frontal or declarative system is demonstrated in Rule-Based tasks, where participants quickly learn explicit rules (once they realize them). In this process, the most important are working memory, in which the testing rule is stored and executive functions, which are responsible for switching once the tested rule proves to be invalid.

Basal ganglia mediated or procedural system is demonstrated in Information Integration tasks. Categories are learned slowly with immediate reinforcement. In this process, connection between visual stimuli and proper motoric reaction is associated through reinforcement.

There is also competition between these two systems. While a declarative system brings results, a procedural system is blocked and vice versa. This model is also well established with cognitive neuroscience data. Each of the systems is precisely described on a neural level, but also, there are many connectionist models which support this theory.

3. THOUGHT AND LANGUAGE

When thought is mentioned, we usually consider internal representation of the outside world along with processes of categorization and reasoning. This relies primarily on semantic representations where the core elements, as it was illustrated in the previous chapter, are concepts. For this reason, when thought is mentioned in the further discussion, it will mostly be related to the concepts.

On one hand, concepts are not the sole property of the human, given that animals are able to form it too. However, this is not the case with language, since homo-sapiens are the only living being which can use symbols. This property made some authors to consider a man as an animal symbolicum (animal that uses symbols), since in it they saw distinctive features of a man compared to other living beings (Cassirer, 1944; Ivić, 1987).

One of the fundamental questions in cognitive psychology and psycholinguistics is the relation between language and thought. This question is also widely known as a *language and thought debate* (Carroll, Von Stutterheim, & Nüse, 2004).

Theories and hypothesis related to this problem vary from the ones which claim that language and thought are completely separated (and that there is supremacy of thought over language) to the others which claim that thought and language are basically the same thing, or at least mutually interdependent and very closely related. Broadly speaking, we can identify two groups of theories: one which claims that there is **cognitive priority** of concepts over language and the others which put forward **linguistic relativity**.

4. COGNITIVE PRIORITY HYPOTHESIS

This group of theories originates in philosophical rationalism and empiricism. While for rationalists, thought was innate, for empiricists, concept of the object should be formed, before we can have a name for it (Locke, [1964 (1690)]; Gleitman & Papafragou, 2005).

Related to the question whether speaking of different languages can make people think differently, the answer that these theorists propose is: no. Thought is first, language is second. For example, German-American anthropologist, Franz Boas, who explored American-Indian languages, realized huge differences in the way these languages express the same thoughts, compared to English and other Indo-European languages (Boas, 1911). However, Boas claimed that these language expressions are expressions of just one part of the complete concepts and that different languages express different parts of the same concepts.

Further support of cognitive priority goes from the famous developmental psychologist, Jean Piaget (Piaget, [1959 (1926)]; Piaget, [1952 (1936)]; Piaget & Inhelder, [2008 (1950)]). Piaget's theory of cognitive development signifies the development of concepts, which goes before language development and is a precondition for it. Only after concepts are developed, we can learn to name them and generally, to talk.

Additionally, work of Noam Chomsky (Chomsky, 1957; Chomsky 1965) gives further support in the disputation of cross linguistic diversities. Chomsky claimed that essential grammar structures of language are innate and that all differences between languages are only the surface differences. This leaves a possibility of various differences between languages, which are not essential and hence, not capable to influence cognition.

At present, cognitive priority hypothesis is primarily a group of similar ideas, rather than a unified theory (Fodor, 1975; Clark & Clark, 1977; Pinker, 1994; Li and Gleitman 2002; Gleitman & Papafragou, 2005; McWhorter 2014). Despite differences, there are some common ideas which could be derived from these writings:

1. Cognition and language are two different and separate processes. As Slobin noted: "Language evokes ideas: it does not represent them. Linguistic expression is thus not a straightforward map of consciousness or thought…" (Slobin, 1979).

2. In relation between the language and cognition, cognition goes first and is a necessary precondition for language. As Gleitman and Paparfagou noted: "*thought is first, language is its expression*" (Gleitman & Papafragou, 2005). Additionally, the role of language is reduced to naming and communication: "*…concepts come first and that language merely names them: nouns name persons, places, or things; verbs name actions and events; adjectives name modifying concepts…*" (Gentner & Goldin-Meadow, 2003).

This means that the conceptual system is completely independent from language. We have concepts, no matter if there is a language or not.

3. Language does not have influence on cognition. Content of concept is not dependant on the existence of labels and language in general. There are not any cognitive consequences of naming and the use of language. No matter if there is a name of the concept or not, concepts will exist unchanged.

Some of the arguments which were used in these writings to support the above mentioned claims were the following:

1. Animals and little babies can think, even though they cannot speak. This proves that cognition is older (phylogenetically and ontogenetically) and consequently independent from language.

2. Language is sketchy, thought is rich (Gleitman & Papafragou, 2005). No matter how rich a vocabulary person has, one cannot express all richness of the thought. In written or spoken language, many things are implicitly considered. This means that we rarely describe all details in spoken communication, which proves broader scope of cognition and cognitive supremacy over language.

3. Language cannot determine perception of space relationship, motion and time (some of the mostly empirically studied areas in linguistic relativity research – presented below). Since there were some empirical findings which were in favour of linguistic relativity, these authors disputed it by interpreting the same results differently. Meaning, they claimed that language differences are just different means used in description of the same thing (for example, Li & Gleitman, 2002).

5. LINGUISTIC RELATIVITY

In the years of development of linguistics as an independent science and when more and more languages were discovered worldwide, the question that was posed was whether these language differences could somehow influence thought? Since there are so many different languages (over 7000 languages in the world) are there any possible cognitive consequences of that?

As it was previously stated, one group of theorists claimed that language can influence cognition and that language and cognition are at least closely related if not completely the same. One of the first who proposed these ideas was German philosopher and linguist, Wilhelm von Humbolt. In his famous book on language, he claimed that language and thought are closely related and since there are so many differences between languages, speakers of different languages perceive and think about the world differently (von Humbolt , [1988 (1836)]). This claim represents the cornerstone of what will become the linguistic relativity hypothesis.

Great influence to this group of theories was given from modern philosophy. Wittgenstein claimed that the "The limits of my language are the limits of my mind" (Wittgenstein, 1922). Furthermore, most of modern philosophy, including schools such us: analytic philosophy, hermeneutics, structuralism, feminism, social constructivism and the entire post-modern philosophy, are based on the preconception that mind and language are closely related and that important philosophical conclusions could be driven from the analysis and study of language (Frege, [1948 (1892)]; Russell, [1996 (1903)]; Russell, [2013 (1943)]; Dilthey et al., 1989; Habermas et al., 2015; Faucault, 2012).

The most popular theory which had a significant influence on the further development of language and thought debate was so called **Sapir-Whorf hypothesis**. Even though never directly specified, it could be considered that this hypothesis was developed by Benjamin Lee Whorf who, apart from his anthropological linguistic studies, also relied on works and views of his professor, Edward Sapir (Sapir, 1929; Whorf 1944; Carroll, 1956). In the essence of this hypothesis is the claim that language crucially influences thought, perception and cognition in general.

Whorf used to work in a fire insurance company and he analysed effects of language on treating objects, which led to the fire starting. For example, only because the kettle was moved off the fire and as something that was "off", it was considered to be safe. However, sometimes it was not and in some cases it led to the fire.

Apart from these nominal effects, this hypothesis considers that grammatical structure of one language constraints and shapes thoughts and perception of the speaker. Meaning that since two languages have a different way of expressing the same thing, speakers of these two languages will have different perception of it. This consequently means, for example, that a Frenchman and a Chinese will have a different perception of the world, since these two languages are nominally, phonologically and primarily grammatically different.

Whorf studied language of Hopi Indians in Arizona. He gives an example of use of plurals in the language of Hopi Indians: unlike Indo-European languages, Hopi Indians have plurals only for visible objects. For example, the concept of several days that are used in Indo-European languages does not exist in this language, since no one ever experienced several days directly. This type of plural in Indo-European languages, Whorf calls "objectified" or "imaginary" (Whorf, 1944). But we do perceive days as quantity in the same manner as we do with objectively quantifying objects (ten men, for example), even if we never saw several days in one place. Additionally, Hopi Indians have different descriptions and consequently perception of time, space and other things. They develop different "habitual thought" than people speaking Indo-European language. This "habitual thought" represents linguistic patterns, rather than simple language and includes a "routine way of attending to objects and events … all give-and-take between the language and the culture as a whole" (Whorf 1944).

We can differentiate two versions of this hypothesis: the **strong** and the **weak** one (Carroll, 2004). According to the strong version, the "presence of linguistic categories creates cognitive categories" (Carroll, 2004, page 382). This means that concepts are essentially created due to the language, and which concepts will be created depends on language, not the outside world. This view is better known as *linguistic determinism*.

On the other hand, a weak version of this hypothesis claims that: the "presence of linguistic categories influences the ease with which various cognitive operations are performed" (Carroll, 2004, page 382). According to this version, language still can have influence on thought, but this influence is not as crucial as in linguistic determinism. Language influences thought by modifying concepts and making some of the cognitive processes easier for the speakers of different languages.

Consequences of these claims were enormously influential. These ideas inspired debate which is still present and many of the current theorists are in favour or against this hypothesis. Additionally, apart from theoretical debate, it influenced many further empirical researches.

Since we can differentiate two ways in which language influences thought: lexical and grammatical influence, we can also identify two different empirical lines of research in which this hypothesis was tested.

As far as *lexical influence* is concerned, vocabulary differentiation across languages was primarily examined. Meaning, it was noticed that some languages have more terms for the same group of objects than others. Frequently quoted is the example of the Eskimo language in which there are around 20 different terms for snow (originally quoted by Boas). More probably, there are only 3 different terms of snow (Brown & Lenneberg, 1954), but it does not change the point that Eskimo language has more terms for snow than most Indo-European languages. The greater number of terms consequently means that Eskimos have 20 (or 3) different concepts of snow, unlike most Americans or Europeans.

Concerning vocabulary differentiation across languages, one of the most popular topics for empirical testing was colour differentiation. Different languages have different number of basic colour terms and this number significantly varies across languages (Berlin and Kay, 1969). For example, Dani tribe in New Guinea has only two terms for colours: dark and light, while most of European languages have around 11 (including English). Eleanor Rosh (previously Heider, 1972) tested whether members of this tribe differentiate focal colours (the most representative hue of one colour) worse than English speakers. She concluded that memory of focal colours of Dani participants was equal to the English speakers, even though their language was poorer in number of language terms for colours.

Research of Eleanor Rosh was considered to be strong proof against the Sapir-Whorf hypothesis. Further researches in this area were abandoned and the focus of research in cognitive psychology was switched from language influence on thought to concepts solely.

Language relativity reappeared in experimental research after almost two decades of hibernation. According to Gentner and Goldin-Meadow (Gentner & Goldin-Meadow, 2003), there were several reasons for this rebirth of language relativity theories.

One of the reasons was further studies of the differences between languages. For example, Choi and Bowerman (Choi & Bowerman 1991) studied spatial terms in English and Korean, and realized that there were significant differences between the two languages in the expression of those terms. These differences consequently lead to a completely different way in which children acquire spatial concepts.

Furthermore, new studies of space perception dependant on language structures appeared, shifting empirical focus from colours (Bowerman, 1996; Brown 1994; Levinson 1996; Levinson, 1997; Levinson, 2003), even though, new researches into colour perception go in favour of the Sapir-Whorf hypothesis (Roberson, Davies & Davidoff, 2000; Winawer et al., 2007).

Finally, one of the reasons was rediscovery of Vygotsky's work (Vygotsky, [1962 (1936)]) who claimed that language can influence and aid cognitive development.

Apart from lexical, in many researches *grammatical influence* on thought was tested. For example, there are studies by Bloom (Bloom, 1981) on the difference in counterfactual reasoning between English and Chinese speakers. Furthermore, studies of grammatical gender expression difference between languages (Martinez and Shatz, 1996; Boroditsky, Schmidt & Phillips, 2003).

In the following decades, numerous researches were conducted in the area of linguistic relativism, which lead to numerous hypotheses that significantly varied. They all had in common that language influence thought, but differed in which way that influence is conducted and how extensive it is.

There are **several classifications** of new hypothesis within *linguistic relativism* (Enfield 2015; Lucy, 1997; Bloom & Keil, 2001), which Enfield named as Neo-Whorfianism (Enfield 2015). However, here will be presented the classification developed by Wolff and Holmes (Wolff & Holmes, 2011). This classification will be presented in a slightly modified order, indicating the continuum from the theories according to which language mostly influences thought to those according to which it least influences thought. This classification includes the following views:

Language-of-thought is a view which claims that *language and thought are the same thing*. Some philosophers like Plato and Kant (according to Wolff & Holmes, 2011) claimed that language and thought are the same thing and that people think with language. This idea was also proposed by some theories in psychology. For example, behaviourists considered thought as its language manifestation (Watson, 1913; Watson, 1920). Since their attention was focused on behaviour, rather than inner processes and introspection, language expression was a kind of a "speaking behaviour". In this sense, thinking is nothing more than internal "silent" speaking. Proof for this, behaviourists found in detected tinny lip movements when a person thinks about something.

Nowadays however, the dominant opinion is that language and thought are two separated things (Murphy, 2002; Casasanto, 2008, Thierry, 2016). Thought could be developed even by little babies and animals, while language is the sole property of the adult homo-sapiens.

Linguistic determinism is previously stated as a strong version of the Sapir-Whorf hypothesis. As it was mentioned, according to linguistic determinism, *language categories do determine conceptual categories*.

This theory, along with the language-of-thought view, is in a sense obsolete and there are not many empirical findings which support them. According to Wollff and Holmes, these two theories could be dismissed, while the following theories are still in use.

Thinking after language considers cognitive effects after the use of language, meaning that language can direct attention to certain elements, which in return can induce some cognitive consequences. There are two views in this group of theories: *language as a spotlight* and *language as an inducer*. In language as a spotlight, certain linguistic properties can make some elements in the outside world more salient and likewise lead to cognitive consequences. In language as an inducer, language can prime the way how information is processed and lead to different processing.

Common characteristic of thinking after language group of theories is that effects of language will not be eliminated, even if there is exclusion of verbal processing.

Thinking with language is based on the mutual processes of language and thought which are activated together. These effects are conducted on-line. The main characteristic of this mutuality is that it can be eliminated by exclusion of verbal processing (by some verbal task or proper electrical stimulation).

We can differentiate two types of theories in this group: the first which proposes that language is a meddler and the second, for which language is an augmenter. *Language as a meddler* is based on mutual concurrency of linguistic and non-linguistic code. Once these codes are consistent (are about the same thing), we have easier processing. In the opposite case, processing is slower.

In the *Language as an augmenter* group, language and thought are combined and likewise, language augments thought. One of the most notable theories from this group is Lupyan's Language Augmented Thought hypothesis (described later).

***Thought before language**: Thinking for speaking* is a theory developed by Slobin (Slobin, 1996; Slobin, 2003). According to this theory, thinking can be shaped by language only in the preparation for language activities (speaking or listening), but that does not influence core cognitive processes.

When it became evident that people use languages that are significantly different and which demand different cognitive resources for its use, cognition was somehow "split" into two types of conceptual representation: *universal representations*, and *representation used for specific language* (Clark, 2003), the so called "thinking for speaking" (Slobin, 1996). Universal representations never get influenced by the specific language, no matter of the structure of the language. On the other hand, thinking for speaking gets influenced by the specific language and can be modified by it.

"Thinking for speaking" does not consider cognitive resources for speaking only, but also for listening, comprehension and formulating utterances. Thinking for speaking represents cognitive resources that were derived from general cognition in order to deal with the specific language. There is nothing in thinking for speaking which does not exist in general cognition. However, what resources will be taken from general cognition depends on the specific language.

In this way, general cognition still remains intact. Any cognitive consequences of use of different languages would be assigned to "thinking for speaking" rather than to general cognition. This construct was used to explain many further arguments attempting to demonstrate that language differences could lead to different cognition. It was also used by theorists who were in favour of cognitive priority hypothesis.

At the end of the elaboration of linguistic determinism, we can discuss some new classifications and new tendencies in this area. For example, we can propose another possible division within the theories related to linguistic relativism: theories that signify *effects of different languages* and theories that signify *effects of language in general*. While the first group of theories consider cognitive differences between speakers of different languages (which is the main line of research which started with the Sapir-Whorf hypothesis), the second group considers effects of language on thought in general, no matter if we speak different languages or not. The concept that is labelled with the name (language) is different from the one that is not. Furthermore, effects of linguistic labels are significantly different compared to other types of labels (non-verbal labels or features).

Regarding some new tendencies, Enfield proposed the new view of linguistic determinism, based on a different view of language functions. While there are six language functions (emotive, poetic, conative, referential, phatic, and meta-lingual) (Jakobson, 1960, according to Enfield 2015), in previous research it used only its referential function. Additionally, concepts are viewed as deprived from its social context. Even though these views are interesting, in this work, a “traditional” view will be used, since more theoretical debates and more empirical findings would be necessary for the Enfield’s proposed view.

In the end it should be noted that while in the past theories of cognitive priority and language relativism were strongly contrasted, modern views are much closer in some elements. Some modern interpretations of cognitive priority admit possibility that language can influence thought (Gleitman & Papafragou, 2013), but deny possibility that these influences alter conceptual representation. According to Gleitman & Papafragou, language can influence thought in two different ways: *directly and permanently*, which is the position of classical Whorfianism, but also in an *indirect and transient* way. The second way is what new interpretation of cognitive priority accepts: language influence thought specifically on-line, but these effects are lost after language is not used. And as it was noted previously, they interpret differently the entire body of research which linguistic relativity theorists conducted, claiming that these effects are not essential and permanent. Like Regier claimed (Regier et al., 2010, page 179), these findings: “*act as a sort of Rorschach test. Those who “want” the Whorf hypothesis to be true can point to the fact that the manipulation clearly implicates language. At the same time, those who “want” the hypothesis to be false can point to how easy it is to eliminate effects of language on perception, and argue on that basis that Whorfian effects are superficial and transient.*” (quoted by Gleitman & Papafragou, 2013, page 518).

These different interpretations of the same empirical findings present huge problems in language and thought debate. It is not always easy to differentiate what are the effects of a “clearly conceptual” or “clearly linguistic” system. This problem also could go in favour of the view that cognitive functions are mutually interdependent and that it is not easy to make strict division between them. The same argument could also stand for the relation between language and thought.

6. EFFECTS OF VERBAL LABELS ON CATEGORISATION

One specific and important problem related to linguistic relativity is effects of verbal labels (naming) on the process of categorisation. In the light of linguistic relativity, the problem could be also further specified as the question: what are the cognitive consequences of naming (Lupyan, 2012a)?

This problem was examined primarily in developmental cognitive psychology, but newer experiments include studies on adults and also use different methodologies, such as ERP and cognitive modelling.

In most of the experiments on children, *dishabituation paradigm* (or novelty preference paradigm) was used (Baillargeon, Spelke & Wasserman, 1985; Schafer & Plunkett, 1998; Schafer, Plunkett & Harris, 1999). This experimental paradigm typically consists of two phases: learning (or habituation) in which children are exposed to the certain stimuli of one category and testing (or dishabituation) in which they are presented with stimuli from the different category. Variations in this paradigm are the preferential looking technique (usually used for very small children) and the touching and examination technique (which is used for slightly older children).

One of the most notable findings of cognitive development theories and its relation to the language was that presence (or learning) of labels in categorisation task facilitates category acquisition (Quinn, Eimas & Rosenkrantz, 1993; Waxman & Markov, 1995; Xu, 2002; Plunket & Hu, 2008; Plunkett, Hu & Cohen, 2008). There are several views of this problem, but most notable are two opposed views, which metaphorically make distinction between: “child-as-theorist” and “child-as-data-analyst” (Waxman & Gelman, 2009). The former resembles Waxman’s *linguistic labels as conceptual marker* view and is rooted in rationalistic tradition, while the latter resembles Sloutsky’s *labels as object features* and is rooted in empirical tradition (Sloutsky & Fisher, 2011).

Labels as conceptual markers

Markman and Hutchinson (Markman & Hutchinson, 1984) discovered that two year old children presented with a task in which they need to select another thing of the same kind to the one presented, typically select the thing which is thematically associated with the target (for example: dog – bone). However, if the children were told that one thing is named in a certain way (for example: “fendle”), and are asked to find another thing named that way, they usually look for a thing that is a member of the same category. This shows that labels point to categories.

Furthermore, **Waxman** (Waxman, 1991; Waxman & Markov, 1995; Balaban & Waxman, 1997; Fulkerson & Waxman, 2007; Waxman & Gelman, 2009) claimed that linguistic labels refer to the concepts. Referring means that “a word links to a conceptual representation that is more abstract than the entities that happen to be present in the naming context” (Lyons, J., 1997 – quoted from Waxman & Gelman, 2009, page 259).

In their studies (for example Waxman & Markov, 1995), they used *novelty-preference paradigm*, where young children (12-13 months) were presented with some objects from the same category (usually toys) and after presented with two objects, one of which was a member of the same category as the previous objects. If a child developed concept, it would show preference to the object from the novel category, rather than the one belonging to the same category it played with before. Once these objects are named (by experimenter), children tend to learn concepts significantly faster (habituate to their presence). In this sense, word serves as an invitation to form a category, even for the objects that are not always perceptually similar (Waxman & Markov, 1995).

Broadly speaking, there is a property of verbal labels that refer to a category. The word is tied to an entire category rather than to the specific object. This tendency to use a label as a *conceptual marker* is somehow innate to children.

In response to the view of labels as object features (presented later), Waxman further specified four crucial assumptions of her hypothesis (Waxman & Gelman, 2009, page 258-259):

1. Words do not merely associate, they refer. This property was already explained.

2. Words and concepts are more than a collection of sensory and/or perceptual features. Very often, words mark concepts which do not have a clear visual counterpart. Some concepts mark absent or abstract things.

3. Words and concepts are not unitary constructs. Words often have different functions (for example: nouns, adjectives, verbs). Different word function mark different concept aspects.

For example, Brown (Brown, 1957) demonstrated that grammatical form of the novel words determines partially their meaning. If the novel word was constructed as a noun (for example "dawe"), it was considered to be an object, if it was constructed as a verb ("dawing") it was considered to be an action. This property was identified for the school children, but even at an earlier age, children are able to make difference between word functions.

4. Words are located within intricate linguistic and social systems. Words are not isolated, but rather belong to the broader context, including linguistic and social. All these contexts are involved in the word meaning which is not explicitly stated in the form of the isolated word alone.

Labels as object features

The model that was developed by **Vladimir Sloutsky** (Sloutsky & Lo, 1999; Sloutsky & Fisher, 2004; Sloutsky & Fisher, 2012; Sloutsky, 2010) explains effects of verbal labels on categorisation treating labels as object features, which are similar to other features, like for example shape and colour. Faced with novel objects, based on the similarities of their features, children induce whether they are members of the same category or not. In that sense, inductive generalization of novel categories is also affected by presence of labels. For example, children will more probably induce that two objects are member of the same category if they share the same name, rather than if they have different or no name at all (Sloutsky, Lo & Fisher, 2001).

This property was further specified and quantified in Slutsky's **SINC** (Similarity Induction Naming Categorisation in young children) model. According to SINC model, as previously stated, labels are treated as features like other features of the objects. When two objects share the same name, this name increases the overall similarity of the objects, which leads to induction that these two objects belong to the same category.

Model calculates similarity between two objects (using multiplicative rule), based on the features that are different. For example, if we calculate similarity between some object B and the target T, similarity can be expressed with the following equation:

$$\text{Sim}(B,T) = S_{v,a..}^{N-k} * S_L^{1-L}$$

Where N is total number of features, k – number of matches, L – label match (label match – L=1, label mismatch – L=0), $S_{v.a.}$ and S_L (ranging from 0 to 1) represent attentional weights of a mismatch (higher number, lower attentional weight) for visual attributes and labels respectively. These attentional weights are not usually the same: they are higher for labels (lower S value), meaning that labels have more influence on similarity compared to visual features.

In most of the experiments that were used in these researches, participants had a task to select one of the two items presented (A or B), compared to the target item (T).This similarity calculation was used to calculate overall probability of selection of one item (let us say, item A). This probability could be calculated with the following equation:

$$P(A) = \frac{Sim(A,T)}{Sim(A,T) + Sim(B,T)} = \frac{1}{1 + \frac{Sim(B,T)}{Sim(A,T)}}$$

Once we use similarity values from the previous equation, we get a result which predicts probability of selection of a certain item.

In one such research (Sloutsky, Lo & Fisher, 2001), Sloutsky demonstrated that his participants (children of 4-5 years) were making selections which highly resembled to the model prediction (R^2=.884). However, this was not the case for older children (11-12 years), who strongly relied on labels. While results between the two groups were more or less comparable in the condition without labels, they highly differed in the label condition.

Sloutsky interpreted these results with developmental differences: while older children were capable to identify that label marks category, younger children do not. For this reason, children use all features (including labels) in estimation of objects similarity.

This model was later updated (Sloutsky & Fisher, 2012) in order to incorporate possible phonological differences between labels. According to an updated model, similarity between two items is calculated using the following equation:

$$\text{Sim (B,T)} = \lambda^{\beta} \upsilon^{B}$$

Where B is an item which is compared to target item – T. λ and υ represent attentional weights of a label and visual property (ranging from 0 to 1), while β and B represent a number of feature mismatches between two labels and appearances of two items (B and T).

This model was tested in several empirical studies (similar to the one presented) in which predictions were confirmed (Sloutsky & Fisher, 2004; Sloutsky & Fisher, 2012). Based on these findings, Sloutsky concludes that labels are *object features* (not conceptual markers) and that they increase overall similarity between items, which leads to induction that two objects belong to the same category.

The advantage of this model is that it uses precise mathematical equations with which it is possible to predict results of categorisation. The problem with this model is that it was developed only for younger children and it is not suitable for adults.

Language augmented thought hypothesis

Both previous theories were developed on children to explain effects of labels on categorisation in developmental context. These authors do sometimes interpret what effects labels can have on adults, but they did not conduct much empirical testing.

Hypothesis that was entirely developed by research conducted on adults is **language augmented thought hypothesis**. This hypothesis was developed by Gary Lupyan (Lupyan, Rakison & McClelland, 2007; Lupyan, 2012a; Lupyan, 2012b; Lupyan, 2015). Lupyan thinks that labels (and language in general) have strong influence on thought. Linguistic labels highlight the most fundamental features of the category. Once labels are learned, they keep making influence on the concept on-line, meaning that activation of the category will influence activation of label. This label will in return activate fundamental features of the concept and further enhance its categorical representation. For the reason of this enhancement of thought, which is actually thought augmented by the language, this theory got the name.

The foundation for his claims, Lupyan got from the work of William James, who claimed that language can influence thought and concept formation. Based on previously stated basic principles, Lupyan predicts further consequences of this view (from Lupyan, 2012a):

1. Verbal labels modulate “non-linguistic” representation,
2. Effects are deep, meaning that they can influence visual processing,
3. Verbal labels are special in the sense that there are no similar effects of other types of labels.

In order to test these predictions, Lupyan conducted series of research. He demonstrated that participants learned novel stimuli (types of aliens) much faster and generalized them better once these aliens are combined with verbal labels than without labels (Lupyan et al., 2007). It is important to signify that these labels were presented to the participants only after their response, so the labels were completely task redundant. In this research, it was also demonstrated that there was superiority of verbal labels compared to non-verbal labels, such as space directions.

In further research, influence of verbal labels on visual perception was demonstrated (Lupyan, 2008, Lupyan, 2010, Lupyan & Spivey, 2010; Lupyan & Thompson-Schill, 2012). For example, in the picture verification task, participants who were presented with the verbal label just before the exposure of the picture, were significantly faster than the participants who were presented with non-linguistic labels, such as sounds related to the picture (for example “moo” sound for the cow). One of the reasons for this effect is identified in more prototypical activation of the concept by the verbal label, than by the non-linguistic label. Basically, verbal and non-linguistic labels activate different elements of the same concept.

In latter work, this effect was interpreted as property of verbal labels as unmotivated cues (Edmiston & Lupyan, 2015). Unlike non-verbal cues which are motivated cues, (such as for example dog barks) which vary with different types of items (bigger dogs bark deeper than small dogs, for example), verbal labels are cues that do not vary with different types of items (dog is a name for all sorts of dog).

On-line effects of language, Lupyan demonstrated with “non-linguistic” tasks using transcranial direct current stimulation - tDCS (Lupyan, Mirman, Hamilton & Thompson-Schill, 2012). Participants showed that cathodal stimulation (inhibitory stimulation) led to poorer performance on the tasks, which was interpreted as quasi impairment produced by blocking of linguistic capabilities. Additionally, it is recorded that aphasic patient have problems with categorisation tasks (Lupyan & Mirman, 2013).

Non-verbal labels. According to our knowledge, there is no systematic research related to the effects of non-verbal labels on category learning and generally, cognitive consequences of non-verbal cues on concepts and cognition. Gary Lupyan demonstrated in several researches (Lupyan et. al., 2007; Lupyan & Thompson-Schill, 2012; Edmiston & Lupyan, 2015) that there is supremacy of linguistic labels over non-linguistic labels (space orientations, object relevant sound, pseudo-sounds). Particularly, there were no same effects of thought augmentation when non-linguistic labels were used instead of the linguistic ones.

Cognitive priority view

This view is not separately developed for the effects of labels on categorisation, but rather it can be extended from the general view according to which language (and also words) do not have any influence on thought. As previously quoted, thought is first, and language is its expression. Existence of concepts is a necessary condition for labels, but labels cannot influence concepts or thought in general.

7. NON-BEHAVIOURAL METHODOLOGY IN LANGUAGE AND THOUGHT RESEARCH

The majority of the previously stated empirical findings and research are based primarily on behavioural methodology. There are however, some limitations of behavioural methodology in the language and thought debate: very often the same behavioural results could be interpreted differently (Gleitman & Papafragou, 2013). Furthermore, it is often impossible to partial out though from the language process. As it was stated previously, it is not always easy to identify independently whether the obtained effects are from thought or language. Furthermore, in non-linguistic tasks, participants usually automatically use language which makes non-linguistic tasks not that non-linguistic (Thierry, 2016).

Finally, the problem is quantity of behavioural experiments. Very often, conducting experiment for each possible experimental manipulation is not practical. Thousands of participants would be needed for research like these.

There are many non-behavioural methodologies used in research related to language and thought debate (and not only for it). The most prominent are *neurobiological methods*, such as: neuropsychological methods (already quoted in the previous chapter - Lupyan's research on aphasic patients), neuroimaging methods (such as ERP, MEG, PET scan, fMRI) and invasive neural methods (tDCS, tACS, TMS, Intracranial stimulation). Additionally, in novel research, the method of *computer stimulation of cognitive processes* is very frequently used, of which the most important is cognitive connectionist modelling.

Finally, another method that is not typical behavioural (but it is) is eye-tracking. This methodology is enormously used in psycholinguistic research, both for children and adults.

These methodologies are constantly growing along with research in which they are used. However, research for each of the listed specific methods will not be presented. Since in this dissertation ERP and cognitive connectionist modelling are used, some of the relevant and representative research (for illustrative purposes – but not an extensive review) in which those methodologies were used will be presented.

Event Related Potential – ERP

This methodology is effective for language and thought research since it is capable to capture cognitive processes at the early stage, while still automatic, before they become conscious and prone to verbalization and various cognitive strategies (Thierry, 2016).

Significant influence related to this methodology in language and thought debate was the work of **Guillaume Thierry** (Thierry, Athanasopoulos, Wiggett, Dering & Kuipers, 2009; Thierry, 2016). Differences in perception that can be produced by different languages, Thierry demonstrated with Greek and English participants whose languages differ in the description of the blue colour (Thierry et. al., 2009). While English have only one colour of blue, Greeks have two: for light and dark blue. Participants needed to respond on squares presented on the screen, but not on the circles. However, circles were painted in light and dark hues of blue and green colours. ERP signal showed differences on early perceptual stages between dark and light blue for Greek, but not for English participants, while there were no differences between light and dark green, for the two languages do not differ in description of this colour (both have only one term for green).

Thierry explains these findings similarly as Lupyan, but unlike Lupyan, he believes that these language effects are not on-line, but rather structural, since such early differences in ERP signal (before 200 ms) are too early to engage any linguistic process.

Similar findings were obtained in object perception, with similar procedures as the previous experiment (Boutonnet, Dering, Vinas-Guasch & Thierry, 2013). English and Spanish participants were examining mugs and cups for which there are two terms in English, but only one in Spanish. Again, there was a difference in early processing in English, but not in Spanish, even though both groups could easily differentiate these two groups of objects.

These findings show that there are language influences on perception even when participants cannot consciously report these processes and when it would be very difficult to examine those effects with behavioural measures. This also implies that successful research of language and thought relation must include multiple methodologies, which would lead to mutually supporting findings.

Connectionist cognitive modelling

Connectionist cognitive modelling is very frequently used methodology in novel research. Sometimes it is much easier to test hypothesis with the connectionist model rather than testing real participants. It specifically includes cases when experimental situations are numerous and it would be impractical to conduct a behavioural experiment. Additionally, connectionist models are biologically plausible (there are neurons and connections in the brain), hence we can presume how the things are organized and implemented in the brain.

There are some doubts related to this methodology, like the one stated by Lupyan related to parameters, that: “the modellers are thought to adjust the parameters to obtain the results they want” (Lupyan, 2012a, page 286). Even thought, connectionist models remain important tools in modern research, which can give important clues of how a cognitive system works and how it is organized.

One of the notable researches in the language and thought debate, specifically related to effects of labels on categorization, which used the connectionist modelling approach, was **Westermann’s** (Westermann & Mareschal, 2014). Westermann tried to integrate most of empirical finding in concept development and effects of labels in concept formation. He developed a neural network in which he included two different neural systems: hippocampal and cortical, the first in charge of quick learning, while the second for slow learning. Stimuli for the network were shared at 4 global level categories composed of 26 basic level categories, of which each had 8 within category exemplars with different feature variations. Two conditions were present: with or without global level category label. Results showed that network formed categories much clearer once global level categories are introduced, which goes in favour of Waxman findings that labels are category markers.

Another connectionist model was proposed by **Gary Lupyan** in which he presented his Language Augmented Thought hypothesis (Lupyan, 2012a). However, this model will be presented in the chapter related to connectionist modelling.

These and other papers showed that connectionist modelling can be useful tool for the exploration of language and thought relations. Furthermore, as it was stated previously, it can be used to describe the way brain really works on a cellular level. And finally, it can be used for the cases when conduction of behavioural experiments would be highly impractical (which will be demonstrated in this dissertation).

8. PROBLEMS, AIM, TASKS AND HYPOTHESES OF THIS DISERTATION

In the previous chapter, when Lupyan's hypothesis of language augmented thought was mentioned, it was stated that he got inspired by the writings of William James. William James, in his "Principles of psychology" stated possibility that language can modulate thoughts during discriminative learning (James, [1931 (1890)]). He gives example of wine tasting: we can have two types of wine on different occasions. These different occasions will drag apart tastes of the wine, helping us to discriminate these two wines (James, [1931 (1890)], page 511). Furthermore, James gives examples of verbal modulation of concept learning: if the "fresh" snow we see in front of our house is named differently, we will form two concepts of snow. This effect is produced since categories of snow are similar, while names are more different. When we tie these names to the categories, these structures become more different than categories without names. So these structures became more discriminable and it is easier to learn them as separated concepts.

A similar explanation is proposed by behaviourist psychologists Miller and Dollard (Miller & Dollard, 1941). They claimed that two similar stimuli, which are associated with two different responses, could increase perceptive differences between them. Since they were behaviourists, spoken words were considered as responses. Consequently, spoken words could influence concept formation.

Before we specify problems of this dissertation, it is important to define the term label. Here, it will be used Lupyan's definition of label as something that is: "a. consistently correlated with a category and b. used to refer to a category" (Lupyan et al., 2007). In this sense, label could be verbal, but also nonverbal, sound, move, spatial relations, actions and similar.

Problems of the dissertation

1. Taking into consideration James' and Miller/Dollard claims, it could be concluded that discriminability between categories, and consequently easiness of their learning, would depend on the level of difference between their labels. Likewise, if there are two categories that are named with phonologically similar words, it would be more difficult to learn them compared to categories which are named with phonologically different words.

This is the first problem that will be analyzed in this dissertation: whether the categories labelled with phonologically different words are easier learned from the categories labelled with phonologically similar words. This problem includes analysis of different level of similarity between verbal labels and its influence on concept formation. Additionally, this includes a comparison of these influences with a situation in which categories are not labelled.

2. The second problem is related to different types of labels. James did not claim that labels need to be verbal. The question is whether different types of labels could also have similar effects on category learning? Specifically, whether non-verbal labels (visual and auditory) with different level of similarity could produce the same effects as stated in the previous problem? Since James never explicitly said that could be the case, we expect that these effects are possible.

3. The third problem is a fusion of the previous two: what is the relationship between effects of verbal and non-verbal labels on category learning? From James' writings, we could conclude that both of them will have similar effects on category learning. However, Waxman and Lupyan thought that verbal labels have a kind of a special status (Waxman & Markov, 1995; Waxman & Gelman 2009; Lupyan et al., 2007; Lupyan, 2012a; Lupyan et. al., 2012; Perry & Lupyan, 2014; Edmiston & Lupyan, 2015) and that there is superiority of verbal labels compared to other types of labels. While Waxman considers labels as conceptual markers, Lupyan considers labels as unmotivated cues.

4. The fourth problem is generated from the analysis of different results in researches related to the effects of labels on category learning. While some of the previously stated papers of Waxman and Lupyan demonstrate these effects, some experiments did not detect it (Batinić, Lalić, Taxitari & Ković, 2015). Some of these researches had different instructions given to participants: some directly asked participants to learn labels, while others did not. Since labels were redundant in most of these experiments for task completion, that could lead to different level of labels learning. Once labels are ignored, there is no effect of the label on an increase of category difference, which could lead to lower performance. Consequently, it might happen that these labels were ignored in the second group of experiments (meaning, not learned) and that was the reason why effects were not identified.

5. The fifth problem is related to creation of cognitive models that could demonstrate obtained results from the previous experiments. Primarily, cognitive models could demonstrate effects of label difference with various difference levels, something which would be very difficult for behavioural experiments. Furthermore, these demonstrations could reveal functional dependence between levels of label differences on category learning.

Additionally, models could demonstrate eventual different effects of verbal and non-verbal labels. And finally, these results could further support and confirm previously obtained behavioural and neural data.

Based on these problems, we can state the aim, tasks and hypothesis of this dissertation.

The aim of the dissertation

The aim of this dissertation is to identify effects of label difference (verbal and non-verbal) on category learning. Additionally, the aim is to identify the relation between these effects, so it could be identified whether there is a special status of verbal labels compared to non-verbal ones. For this purpose, a multi-methodological approach will be used: behavioural – with measures of accuracy and reaction times, neuro-physiological – with measures of event related potentials (ERP) and cognitive modelling approach.

Tasks of the dissertation

Based on the problems and the main aim of the dissertation, the following tasks could be stated:

1. Identify effects of the level of phonological differences of verbal labels on category learning, using behavioural and neuro-physiological measures.
2. Identify effects of the level of differences of non-verbal labels on category learning, using behavioural and neuro-physiological measures.
3. Identify relations between the effects of verbal and non-verbal labels on category learning, using behavioural and neuro-physiological measures.
4. Identify the effects of experimental instruction on category learning, using behavioural measures.
5. Create physiologically plausible cognitive models which will demonstrate the effects of label differences on category learning.

The entire structure of the dissertation could be presented with the following table:

Table 0-1: Factorial structure of the dissertation

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Modality</td></tr>
<tr><td>Auditory</td><td>Visual</td></tr>
<tr><td rowspan="2">Linguistic</td><td>Verbal</td><td>Minimal vs. Maximal</td><td>Minimal vs. Maximal</td></tr>
<tr><td>Non-verbal</td><td>Minimal vs. Maximal</td><td>Minimal vs. Maximal</td></tr>
</table>

An additional factor that was included in the fourth aim presents all these factors with the different instructions participants received in the experiment.

General hypotheses of the dissertation

Within each of previously stated tasks it could be identified general and specific expectations – hypotheses. Here will be specified general hypotheses of the dissertation, while each of the following chapters will further specify and define specific hypothesis of each chapter.

We can identify the following general hypotheses:

1. Learning of categories labelled with phonologically different verbal labels will be faster and easier generalized compared to the categories labelled with phonologically similar labels.

2. Learning of categories labelled with phonologically different non-verbal labels will be faster and easier generalized compared to the categories labelled with phonologically similar labels.

3. Linguistic and non-linguistic labels (visual or auditory) have the same effects on category learning.

4. Experimental instructions in which it is explicitly required from the participant to learn the label will have a higher effect on category learning, compared to the experiment in which there are no experimental instructions.

5. It is possible to create physiologically plausible cognitive models which will demonstrate effects of label difference on category learning.

The following chapters will in detail specify and test each of the tasks and hypotheses in greater detail (each chapter per task/general hypothesis).

CHAPTER I

EFFECTS OF VERBAL LABEL DIFFERENCES ON CATEGORY LEARNING[1]

[1] Adapted version of this chapter was submitted for publication

1. INTRODUCTION

As it was stated in the introductory chapter, we represent the world around us by the use of cognitive concepts. Concepts are representations of entities from reality which are grouped in broader categories (Murphy, 2002; Sloutsky, 2010). This ability is not the exclusive property of a human, but also animals can create concepts. Language, on the other hand is exclusive property of homo-sapiens. Human ability to use symbols made some authors to consider a man an animal symbolicum (Cassirer, 1944).

The relation between concepts and language is the object of debate between philosophers and psychologists for centuries. Roughly, ideas regarding these relations could be split into two broad and more or less coherent groups.

The first group of authors consider language and concepts strictly separated (Fodor 1975; Li & Gleitman, 2002; Gleitman & Papafragou, 2005; Gleitman & Papafragou, 2013; Li, Dunham & Carey, 2009; Klemfuss, Prinzmetal & Ivry, 2012). There is a primacy of concepts over language and concepts exist independently of language. Language is used strictly for communication and cannot influence the content of the concepts. As Gleitman and Papafragou noted: "*thought is first, language is its expression*" (Gleitman & Papafragou, 2005, page 634). Some authors (Lupyan, 2012a) name this group of theories "cognitive priority hypothesis".

The second group of theories consider language and concepts interactive and mutually dependant. Words stabilize complex ideas in the working memory and also influence the content of the concept (James, [1931 (1890)]; Vygotsky, 1962; Lupyan, 2012a). This group of theories is identified as *linguistic relativity hypothesis*, which is brought to its extreme in what is known as the Sapir-Whorf hypothesis (Whorf, 1941; Whorf, 1956). This hypothesis considers specific language as a determinant of mind (for that reason it is named as linguistic determinism). Words we use in specific language and more importantly the grammar structure of a language determines the way we think and perceive outside world.

The probable mutual influence between language and categories is empirically tested and further specified in Lupyan's *Language-augmented thought hypothesis* (Lupyan et al., 2007; Lupyan, 2012a; Lupyan et al., 2012). According to this hypothesis, names are not only used for communication and labelling the concepts, but they also modulate object representation on-line. The relationship between words and concepts is bidirectional.

Lupyan tested his hypothesis in experiments in which participants learned labelled and non-labelled novel categories (Lupyan et al., 2007). Results showed that labelled categories were learned faster and generalized better than non-labelled ones.

The way, in which language can influence the concepts, William James tried to explain in his "Principles of psychology", specifically in the chapter related to discrimination and comparison (James, [1931 (1890)], pages 508-515). James claimed that there are at least two distinct causes which lead to improved discrimination: "*First, the terms whose difference comes to be felt contract disparate associates and these help to drag them apart. Second, the difference reminds us of larger differences of the same sort, and these help us to notice it*." (James, [1931 (1890)], page 510).

Accordingly, the reason for discrimination of concepts lies in the more easily discriminated experiences associated with those concepts. James gives an example of discrimination between two wines: the flavour of the wines we associate to the situations where we tasted them. Since these situations are different, they will drag apart flavours of the wines too. Formally, if concepts B and C are difficult to discriminate, but A and D could be easily discriminated, if A is adhered with B and D is adhered with C, AB and CD will be more easily discriminated then B and C solely (James, [1931 (1890)], page 511). This case could be presented on the following figure (Figure I-1), where distance between items on a horizontal axis represents a level of difference between them.

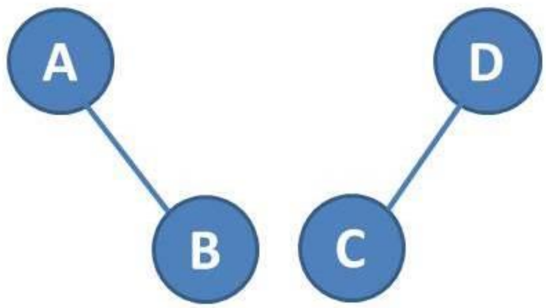

Figure I-1: Model of James' hypothesis

In this case, A and D could be previous experiences related to the concepts. Apart from experiences, James claims that role of A and D could be taken by words, or more specifically names of concepts. Since "*the names differ far more than the flavours*" (meaning wine flavours), "*and help to stretch these latter farther apart*".

In line with aims stated in the introductory chapter, the aim of this chapter is to test whether learning of categories labelled with phonologically different verbal labels will be faster and more easily generalized compared to the categories labelled with phonologically similar labels. This is in line with the previously stated James' hypothesis related to verbal labels: If distinctive elements (names) adhered with two concepts are contributing distinction to these concepts, then if those elements are more distinct we could expect that distinction will be higher between them. Meaning, categories labelled with the phonologically more distinct words would be learned easier compared to the same categories labelled by the less distinct words.

Additionally, if James' hypothesis was entirely correct, we could expect that continuum of label difference leads to a continuum of effects on concept learning and generalization. Since it is enormously difficult to test behaviourally all possible variations of the label differences, we could use a level of phonological difference that lies between maximal and minimal difference, which we can call middle difference. If the results show that concepts that are named with middle different labels are learned and generalized faster than those named with minimal difference, but slower than those with maximal difference, we could conclude that there is a continuum between the level of learning compared to the level of label difference.

Finally, the effect of different kinds of label pairs compared to the no label condition could be tested. In this way we can demonstrate the effects of different features of labels on category learning. It is expected that condition with maximally different labels will be learned faster and generalized better than a silent condition, as in the Lupyan's experiment (Lupyan et al, 2007). Furthermore, we expect that this effect will be lower for minimally different condition.

For this purpose, experiment with categorization task was designed in which participants studied new concepts. The design and materials were similar to those used by Lupyan (Lupyan et al., 2007), except that phonological features of labels were manipulated.

Additionally, electro-physiological measures are included (event-related potentials) which are used to measure brain responses in each of the experiment phases (learning, generalization and testing). It is expected that the cognitive load in the first half of the learning phase will be higher compared to the second half, manifested in higher P300 amplitude (Luck, 2014; Polich & Kok, 1995).

P300 component is sensitive to cognitive load related to the task difficulty. Furthermore, the first half of the learning phase requires higher effort from the participants compared to the second half (since stimuli are completely novel anymore in the beginning). For that reason, smaller amplitude of P300 component is expected in the second part compared to the first part of the learning phase.

Furthermore, it is expected that in the label test phase, there will be no difference in semantic expectances between minimal and maximal conditions, which will be manifested in the similar amplitude of N400 component (Luck, 2014; Kutas & Federmeier, 2011). Given that N400 is sensitive to semantic congruencies, once both types of the labels are learned to the equal extend, it is expected that there will be no differences in the level of amplitude of this component.

2. EXPERIMENT 1

In this experiment, the effects of phonologically minimally and maximally different labels on category learning were tested. As additional conditions, effects of middle different and no label (silent) conditions were used.

Method

Participants

The participants were 80 psychology students, who participated in the experiment as part of their course credit. Participants were randomly divided into four groups. In the first group, participants were presented with *minimally*, in the second group with *maximally* and in the third with *middle different* labels. In the fourth group, participants did not hear any label (*silent condition*). One participant was excluded from the further analysis, since she incorrectly understood the instructions. Additionally, results for one participant in the label test phase for the silent condition were excluded, for the same reason.

Material

The experiment was designed in the program SuperLab 4.0. The screen was a 21" LCD, with a visual angle of 12.23° (distance 70cm). Participants used keyboard buttons which were marked with the arrows (directing left and right) for response. Each participant received auditory stimuli through the headphones. All auditory stimuli were digitally recorded.

During the entire experiment, EEG brain waves of the participants were recorded. This EEG signal was recorded in uni-polar design through electrodes placed at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, CP5, CP6, T3, T4, O1 and O2 locations in line with International 10-20 standard (Luck, 2014) which were showed in Figure I-2. The ground electrode was placed at the centre of the forehead, while the referent electrodes were placed at the earlobes of each participant (A1 and A2).

"PSYLAB EEG8 biological amplifier" in combination with "PSYLAB SAM unit" (Contact Precision Instruments, London, UK) was used for EEG recording. Impedance between the skin and each of the electrodes was below 5kΩ. During recording, the signal was amplified to 20k and filtered with band-pass filters, so the signal in the range of 0.03Hz – 40Hz was recorded. Sampling frequency was 500Hz.

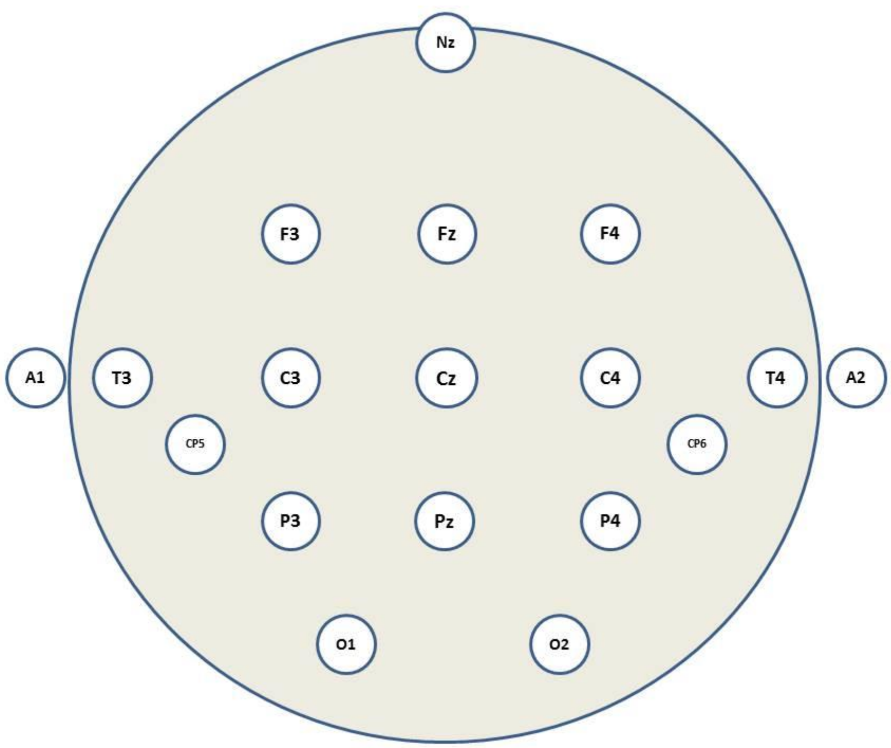

Figure I-2: Location of the electrodes

Stimuli. YUFO stimuli list of aliens was used (Gauthier, James, Curby & Tarr, 2003), divided into two groups (Figure I-3). One group of stimuli had a rounded basis and rounder heads.

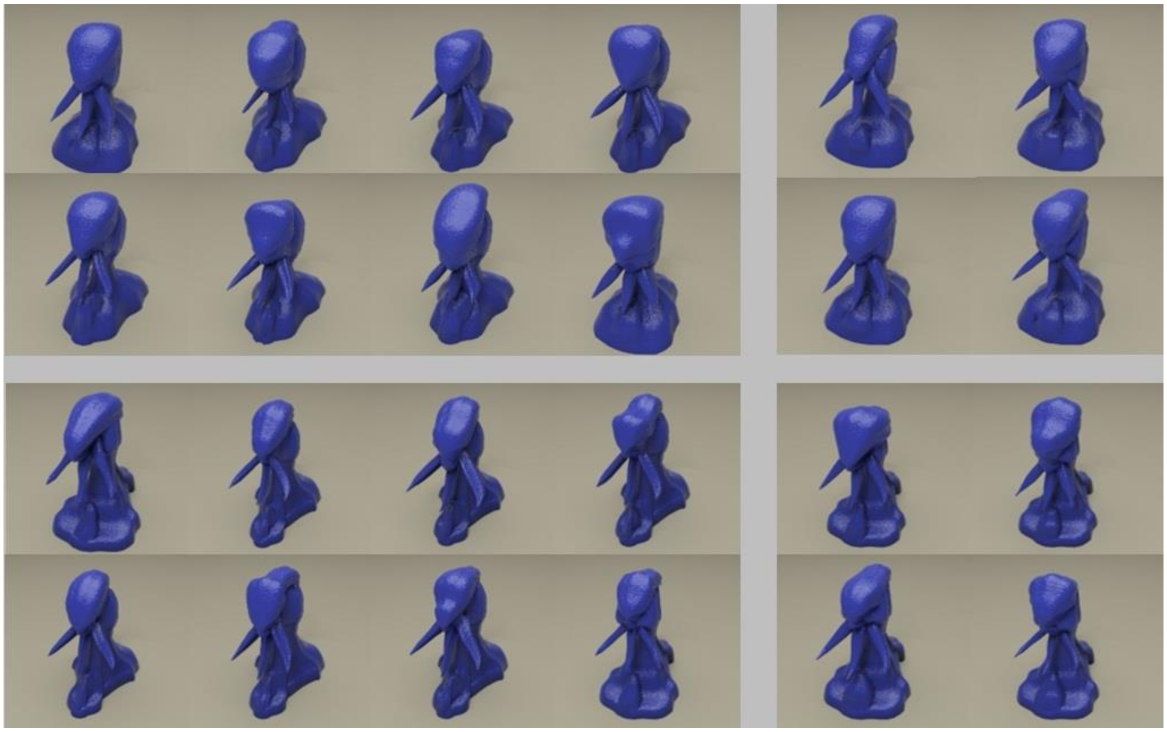

Figure I-3: YUFO stimuli divided into two categories

For the labels, we used pseudo nouns with a difference manipulated on the three dimensions: phonological structure of the label (CVCVC versus VCVC), sonority gradient for alveolar/postalveolar sounds and vowel position (/i/ - /e/ - /a/ - /o/ - /u/). An additional dimension of difference between labels was sound symbolism. Previous research showed that some pseudo words are characterized as more "rounded" and some others as more "spikey" (Köhler, 1929/1947; Ković, Sučević & Styles, 2017). Even though we do not relate these labels to "rounded" or "spikey" objects, we expect that difference based on this property could increase overall difference between labels and maximize it further.

Phonologically minimally different labels differed in the identity of only a single phoneme (consonant in our case) on a minimal scale (džoset (/dʒoset/[2]) vs đoset (/dʑoset/)). Phonologically maximally different labels differed in the identity of all three dimensions in a maximum scale (ketsi (/ketsi/) vs ubom (/ubom/)). Additionally, these labels differed on the dimension of sound symbolism (*ketsi* more "spikey" and *ubom* more "rounded").

Phonologically middle different labels (đosiš (/dʑosiʃ/) – đuzetč (/dʑuzetʃ/)) had the same phonological structure (CVCVC), but differed on both the sonority gradient and used vowels. This difference was minimal for both of these dimensions, which makes a difference between these two labels somewhere in the middle. In the silent condition, the labels were not presented.

Labels were digitally recorded (female voice) of equal length (700ms).

[2] All labels are spelled in brackets according to IPA (International Phonetic Alphabet)

Design and procedure

The experiment was divided into three phases: *learning phase*, *concepts test phase* and *labels test phase*. Participants were instructed that they will do a categorization task in which they will take a position of an explorer discovering a new planet. There are two types of creatures on the planet: good and bad. If they think that a creature showed on the screen is good, they should approach it by pressing a button directing towards the creature. Otherwise, they should escape from it by pressing a button directing away from the creature. In the instruction it was stated that previous missions named these types of creatures by different labels, which would be presented after they receive feedback. Participants were instructed to pay special attention to these names, since their knowledge of it would be tested at the end of the experiment.

In the first, *learning phase* (Figure I-4), pictures of the aliens were randomly presented in the middle of the screen, followed by a picture of the cosmonaut, which was shown on one of the randomly selected (left or right) sides of the screen. Depending on the position of the cosmonaut, participants needed to press appropriate arrow (left or right) in order to approach or move away from the alien. After the button was pressed, participants received feedback (a bell or a buzz), which signalized if their selection was correct. Finally, participants heard labels of the category (names of the group of aliens). It is important to notice, that these labels were completely task redundant, since participants did not need these labels to solve the problem itself.

The learning phase consisted of nine blocks (in each block the entire list of stimuli was presented). Additionally, in this phase, brain potentials were recorded from the onset of the picture of the alien on the screen.

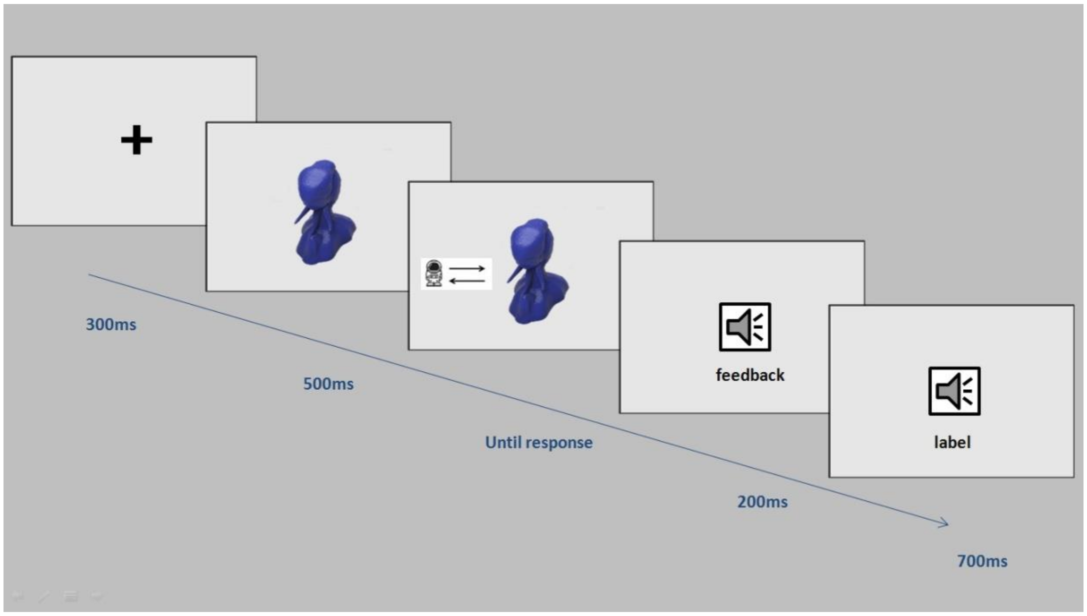

Figure I-4: Trial structure in the learning phase (phase I)

The *concepts test phase* was the same as the learning phase, except that feedback and labels were not presented. Furthermore, in the test phase additional stimuli of the same categories were presented (four by each category, presented on the right side of the grey line in Figure I-3), that were not presented in the learning phase. This was used to test whether participants learned types or specific tokens of the categories. This phase consisted of four blocks. Brain potentials were also recorded from the onset of the picture of the alien on the screen.

In the *labels test phase*, participant had a task to confirm (by pressing the appropriate keyboard button) whether the image of an alien was congruent with the previously presented label (Figure I-5). All the stimuli from the first phase were presented twice in each block (32 trials in total), each time paired with the correct or incorrect label (half of the time correct). There were three blocks in the entire phase.

During this process, the brain potentials were recorded from the onset of the picture of the alien on the screen. Participants responded only after the response panel (with a plus and minus sign) was presented on the screen, by pressing the button on the same side as the plus sign if the name and the alien picture were congruent (and vice versa). The distribution of plus/minus position on the screen was randomly selected. In this way, any lateralized readiness potential (LRP) that could have distorted ERP responses were prevented, since participants did not know on which side of the screen the plus (or minus) sign will be presented.

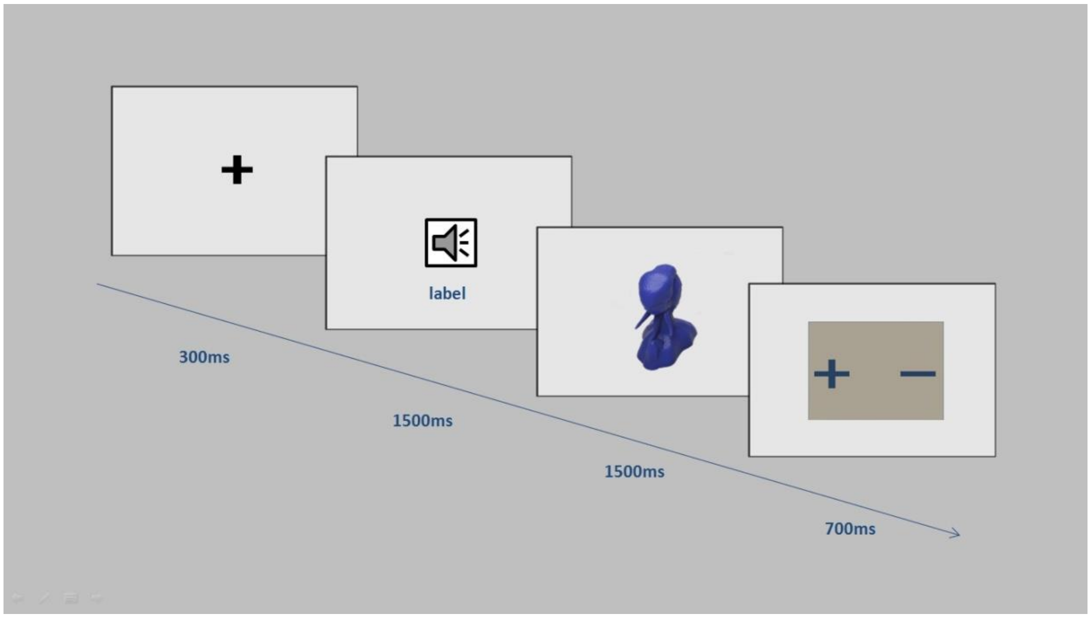

Figure I-5: Trial structure in the label the test phase (phase III)

However, this phase was different for the silent condition, since there were no labels in it. Instead of judging over congruency, participants had a task to identify a good or a bad alien. They did it by pressing the button which was on the side where the plus sign was if they thought the alien was the good or minus side if they thought the alien was the bad. Signs were also randomly distributed on both sides.

As a dependent variable, the percentage of correct responses was measured (accuracy). Furthermore, as a control variable, reaction time was recorded. Finally, an ERP measure of voltage was used, expressed in micro volts.

Results and discussion

Behavioural results

Descriptive results in *Reaction time analysis* for each of the phases are presented in the following table (Table I-1):

Table I-1: Descriptive results of reaction times (in milliseconds) in Experiment 1

Group		**N**	**Mean**	**St.Deviation**
Training	Min.diff	19	1033.99	451.12
	Max.diff	20	872.12	229.08
	Mid.diff	20	1073.84	399.26
	Silent	20	1023.59	310.88
Test	Min.diff	19	1036.55	488.18
	Max.diff	20	768.41	173.56
	Mid.diff	20	995.34	396.82
	Silent	20	882.12	288.55
Label_test	Min.diff	19	800.93	180.16
	Max.diff	20	659.24	142.37
	Mid.diff	20	850.59	282.77
	Silent	20	654.66	167.99
TOTAL	Min.diff	19	957.16	334.37
	Max.diff	20	766.59	136.65
	Mid.diff	20	973.25	298.9
	Silent	20	853.45	237.38

To analyse these results, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor and Phase as a within-subjects factor with two levels (traning and test).

There was no significant main effect of Group ($F(3,75) = 1.913$, $p = .135$, $\eta^2 = .071$). However, there was a significant main effect of Phase ($F(1,75) = 6.347$, $p = .014$, $\eta^2 = .078$). Additionally, there was no significant Phase x Group interaction ($F(3,75) = .899$, $p = .446$, $\eta^2 = .035$).

These results showed that participants faster responded in the test than in the training phase, which is expected, since they already knew most of the stimuli in the test condition and took them less time to make a selection. Furthermore, participants did not differ between groups, so there was no speed-accuracy trade off.

Accuracy. Descriptive results for accuracy (percentage of correct responses) are presented in the following table and chart (Table I-2 and Figure I-6):

Table I-2: Percentage of correct responses in Experiment 1

Group		N	Mean	St.Deviation
Training	Min.diff	19	71.71	12.62
	Max.diff	20	82.50	7.27
	Mid.diff	20	76.84	10.44
	Silent	20	73.33	11.51
Test	Min.diff	19	79.33	12.22
	Max.diff	20	85.47	5.26
	Mid.diff	20	82.19	11.77
	Silent	20	74.43	16.40
Label_test	Min.diff	19	65.57	14.88
	Max.diff	20	79.06	12.16
	Mid.diff	19	70.89	17.05

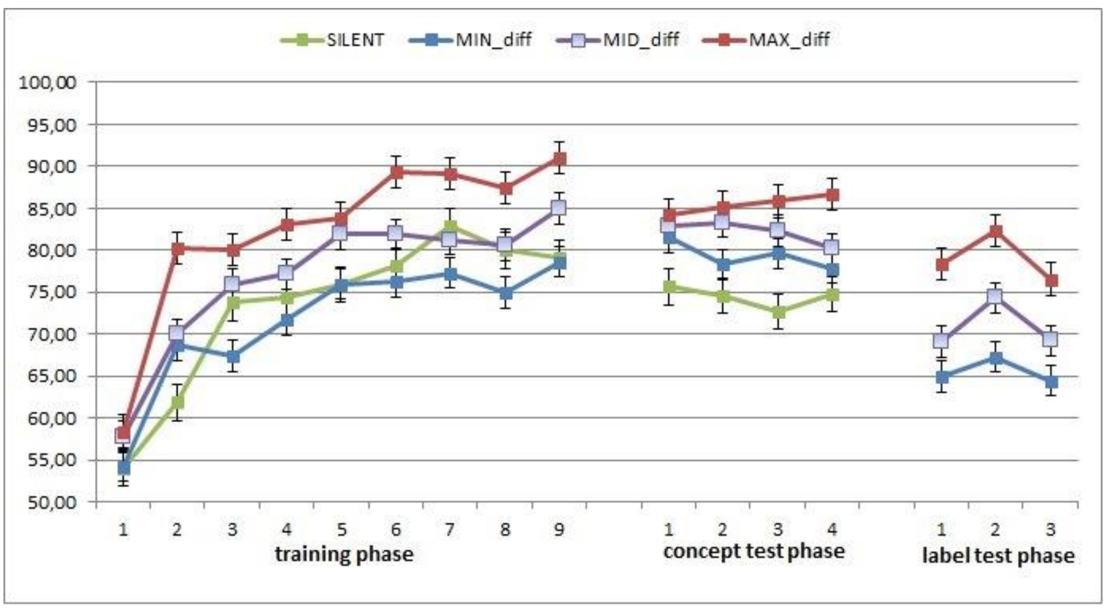

Figure I-6: Percentage of correct responses over phases and blocks in Experiment 1

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor and Phase as a within-subjects factor with two levels (training and test).

There was a significant main effect of Group ($F(3,75) = 3.411$, $p = .022$, $\eta^2 = .120$). Furthermore, there was a significant main effect of Phase ($F(1,75) = 32.849$, $p < .001$, $\eta^2 = .305$). Finally, there was a significant Group x Phase interaction ($F(3,75) = 3.610$, $p = .017$, $\eta^2 = .305$).

Post-hoc comparison of between-group differences showed that there is a significant difference between maximal compared to minimal (mean-difference = 8.464, $p = .018$) and silent condition (mean-difference = 10.105, $p = .04$). Other groups did not differ significantly. It is interesting to notice that the middle different condition was between minimal and maximal levels, but was not significantly different from either of them.

For further analysis of differences between groups, a one-way ANOVA was used for each of the experimental phases, with a Group as a between-subjects factor. There were significant differences between groups in the learning phase ($F(3,75) = 3.987$, $p = .011$, $\eta^2 = .138$) and in the test phase ($F(3,75) = 2.996$, $p= .036$, $\eta^2 = .107$). Since there were no labels in the silent condition in the label test phase, this group was excluded from further analysis. There were also significant differences between groups in label test phase ($F(2,57) = 4.130$, $p = .021$, $\eta^2 = .131$).

Obtained within-subjects differences are due to the learning. It is expected that participants will perform better in the test phase when they already formed the concepts than in the entire training phase.

Interaction between Phase and Group was due to the difference between the silent condition in training and test phase, meaning that the pattern of this condition was different then for the rest of the groups. This gets obvious if we exclude the silent condition from the analysis, interaction becomes insignificant. Interpretation of this result is that the silent condition is not that well generalized as other conditions, which we will further test in post-hoc analysis.

Post-hoc comparison for the learning phase showed differences between maximal different compared to minimal (mean-difference = 10.790, $p = .002$) and silent condition (mean-difference = 9.167, $p = .008$), while other differences were not significant. In the test phase, the silent condition was significantly different compared to the maximal (mean-difference = -11.042, $p = .005$) and middle different condition (mean-difference = -7.761, $p = .046$). Finally, in the label test phase, significant differences were only between the minimal and the maximal difference condition (mean-difference = -13.492, $p = .006$), while the middle different condition did not differ from any of the other two.

These results go in favour of James' hypothesis. The obtained differences demonstrate that participants learned faster and generalized better in maximal difference condition compared to minimally different condition. Additionally, since the middle difference group lies between minimal and maximal difference levels, we can conclude that there is a continuum of effects of label difference on category learning, as we predicted in the introduction. Finally, data revealed that even though there are no significant effects of minimally and middle different labels on category learning compared to the silent condition in the training phase, their effects are obvious in the test phase for middle difference and in the presence of interaction between minimal and silent condition. This goes in line with Lupyan's hypothesis that labels affect generalization, since they are connected to the typical features of the category, no matter if they are similar or dissimilar.

3. EXPERIMENT 2

In experiment 1 we obtained results which go in favour of James' hypothesis. Hoverer, alternative interpretation of the obtained effect of maximally different labels is possible and it could be explained by sound symbolism. In some previous research (Kovic, Plunkett & Westermann, 2010; Sučević et al., 2015), it is showed that sound symbolism can have effects on concepts and its discrimination. Apart from being phonologically maximally different, the labels we used in experiment 1 were also different in sound symbolism.

The question is what effects will produce labels in which there is no sound symbolism, but there is maximal phonological difference between labels? If these labels still produce effects presented in the previous experiment, we can conclude that the effects we obtained are due to the phonological difference, rather than sound symbolism solely.

Additional problems present sound symbolic congruency with stimuli presented. One previous research (Lupyan & Casasanto, 2014) in which the same stimuli as in this experiment were used, showed effects of sound symbolic congruence ("spikey" label with "spikey" aliens vs "spikey" label with "rounded" aliens). Meaning, participants learned faster and generalized better a congruent combination between label and stimuli compared to incongruent. Within this research, a norming study was conducted which showed that participants classified the first group of aliens as "rounded" (top two rows in Figure I-3) and the second group of aliens as "spikey" (lower two rows in Figure I-3).

In experiment 1, labels differed on the dimension of sound symbolism, but the "spikey" label (ketsi) was related to a "rounded" object and vice versa. In that experiment, sound symbolism was used only as a factor that could increase the overall difference between labels.

Based on previous statements, the aim of this experiment is to test what are the effects of sound symbolism on category learning compared to the effects of labels without sound symbolism. For this purpose, we designed the experiment in which we used maximally different words with congruent sound symbolism (*ketsi* with "spiky" aliens) and maximally different words without sound symbolism. With this experiment, we wish to measure the effects of sound symbolism on category learning, while other phonological differences are maximized.

This extension of the previous experiment will be additionally contrasted and compared with the results obtained in experiment 1 (with minimal and maximal difference condition). In this way we can test whether the obtained effects of label difference in experiment 1 could be assigned to the sound symbolism, or they are the sole property of the phonological difference between labels. We expected that the effects obtained in previous experiment could be assigned to phonological difference, rather than to sound symbolism, even though the effects of sound symbolism could be also identified.

Method

Participants

The participants were 40 psychology students, who participated in the experiment as part of their course credit. Participants were randomly divided into two groups. In the first group, participants were presented with maximally different verbal labels without sound symbolism, while in the second group with maximally different labels with sound symbolism which was congruent with visual stimuli (aliens). Results for one participant were excluded from the further analysis, since she incorrectly understood the instructions.

Material and stimuli

Materials and stimuli were completely the same as in the previous experiment, except that the additional labels were used for maximally different labels without sound symbolism. These labels (citech (/citetʃ/) – nudžoz (/nudʒoz/)) differed on the same three dimensions as other maximally different labels, but without sound symbolism.

In the maximally different labels which were congruent with stimuli, same labels as in experiment 1 were used (ketsi-ubom).

Design and procedure

Design and procedure were the same as in the previous experiment.

Results and discussion

Behavioural results

In *Reaction time analysis* for each of the phases the following descriptive results were obtained (Table I-3):

Table I-3: Descriptive results of reaction times (in milliseconds) in Experiment 2

Group		N	Mean	St.Deviation
Training	Max.diff no ss	20	958.86	399.22
	Max.diff congr	19	981.98	465.16
Test	Min.diff no ss	20	902.73	333.04
	Max.diff congr	19	836.15	282.77
Label_test	Min.diff no ss	20	811.86	220.41
	Max.diff congr	19	788.34	272.94

To analyse these results, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor and Phase as a within-subjects factor with two levels (training and test).

There was no significant main effect of Group ($F(1,37) = 0.036$, $p = .850$, $\eta^2 = .001$). However, there was a significant main effect of Phase ($F(1,37) = 6.841$, $p = .013$, $\eta^2 = .156$). Additionally, there was no significant Phase x Group interaction ($F(1,37) = 1.349$, $p = .253$, $\eta^2 = .035$).

As in the previous experiment, results showed that participants faster responded in the test than in the training phase, which is expected, since they already knew most of the stimuli in the test condition and took them less time to make a selection. However, participants did not differ between groups, so there was no speed-accuracy trade off.

Accuracy. Descriptive results for accuracy (percentage of correct responses) are presented in the following table and chart (for each phase and within phase blocks):

Table I-4: Percentage of correct responses in Experiment 2

Group		N	Mean	St.Deviation
Training	Max.diff no ss	20	80.49	10.61
	Max.diff congr	19	81.25	5.27
Test	Min.diff no ss	20	85.42	7.81
	Max.diff congr	19	85.91	5.92
Label_test	Min.diff no ss	20	78.28	17.73
	Max.diff congr	19	78.24	10.87

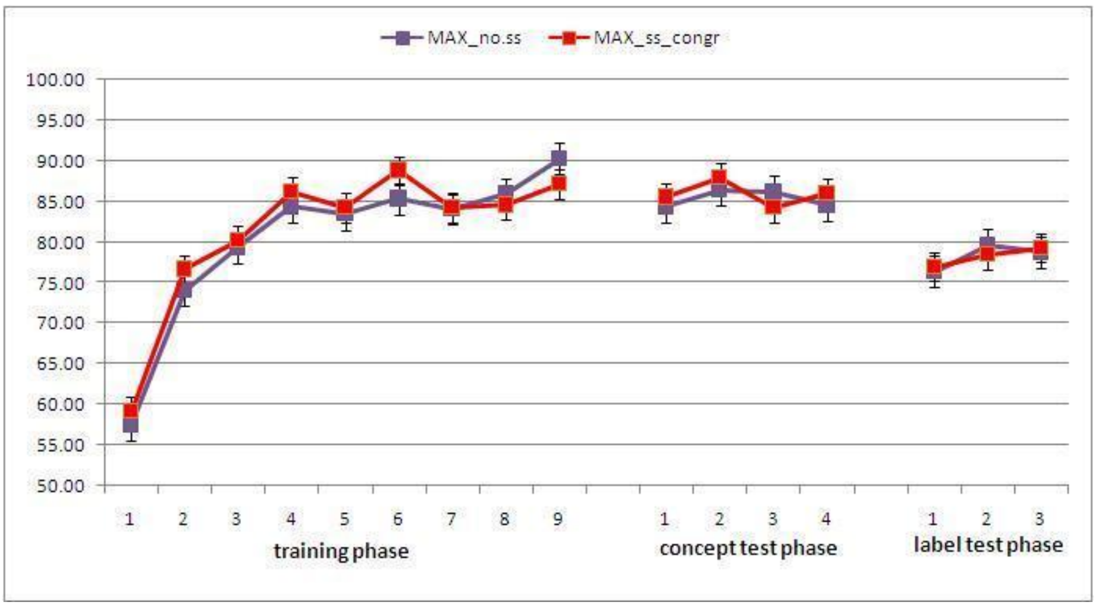

Figure I-7: Percentage of correct responses over phases and blocks in Experiment 2

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor and Phase as a within-subjects factor with two levels (training and test).

There was no significant main effect of Group ($F(1,37) = 0.079$, $p = .780$, $\eta^2 = .002$). Furthermore, there was a significant main effect of Phase ($F(1,37) = 19.676$, $p < .001$, $\eta^2 = .347$). Finally, there was no significant Group x Phase interaction ($F(1,37) = 0.016$, $p = .901$, $\eta^2 = .000$).

In the label test phase, there was no significant difference between the two groups ($F(1,37) = 0.391$, $p = .536$, $\eta^2 = .010$).

There are no differences between the two groups, except in Phase factor, which was expected, due to the learning.

4. COMPARISON OF RESULTS IN EXPERIMENT 1 AND EXPERIMENT 2

As we previously stated, the results from this experiment were compared with the results for two groups from the previous experiment (the maximal difference with incongruent sound symbolism and a minimal difference condition).

Behavioural results

Reaction time. A two-way mixed-design ANOVA was conducted with Group as a between-subjects factor (four levels) and Phase as a within-subjects factor with two levels (training and test).

There was no significant main effect of Group ($F(3,74) = 1.302$, $p = .280$, $\eta^2 = .050$). However, there was a significant main effect of Phase ($F(1,75) = 6.048$, $p = .016$, $\eta^2 = .076$). Additionally, there was no significant Phase x Group interaction ($F(3,74) = 1.050$, $p = .376$, $\eta^2 = .041$).

Results showed that participants faster responded in the test than in the training phase, which is expected, since they already knew most of the stimuli in the test condition and took them less time to make selection. Furthermore, participants did not differ between groups, so there was no speed-accuracy trade off.

Accuracy. Descriptive results for all four experimental conditions are presented in the following plot:

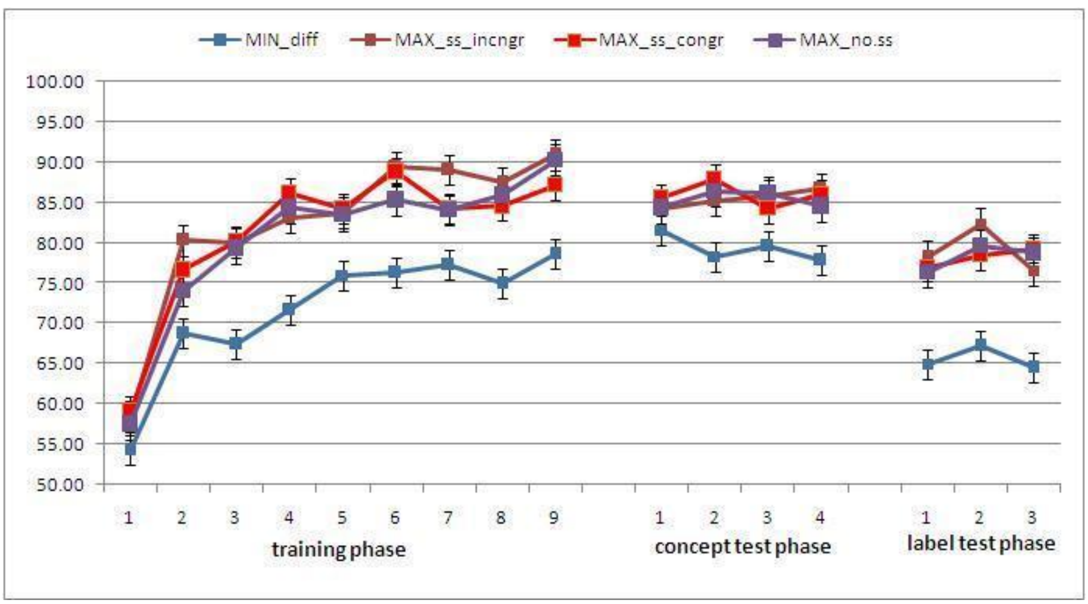

Figure I-8: Percentage of correct responses over phases and blocks

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor and Phase as a within-subjects factor with two levels (training and test).

There was a significant main effect of Group ($F(3,74) = 5.595$, $p = .005$, $\eta^2 = .157$). Furthermore, there was a significant main effect of Phase ($F(1,74) = 46.921$, $p < .001$, $\eta^2 = .388$). Finally, there was no significant Group x Phase interaction ($F(3,74) = 1.697$, $p = .175$, $\eta^2 = .064$).

Post-hoc comparison of between-group differences showed that there is a significant difference between minimal compared to maximal difference without sound symbolism (mean-difference = -7.431, p = .006), maximal difference with incongruent sound symbolism (mean-difference = -8.464, p = .002) and maximal difference with congruent sound symbolism (mean-difference = -8.059, p = .003). Other groups (all maximal groups) did not differ significantly.

As in the previous experiment, within-subjects differences are due to the learning. It is expected that participants will perform better in the test phase when they already formed the concepts than in the entire training phase.

From these results, we can conclude that there is a unique effect of maximal phonological difference on category learning which is independent from sound symbolism. The reason why differences in congruent sound symbolic condition were not obtained (as in Lupyan's experiment) is probably because of the ceiling effect: participants' performance was around 90% on average in all maximally different conditions. Additionally, in Lupyan's experiment, there were sixteen blocks of learning (compared to the thirteen we have here) and still the significant level was p = 0.04. Finally, prior to the experiment, Lupyan conducted a norming study, not relying only on sound symbolism theory.

ERP results

Apart from behavioural, data from event-related potential were also collected. Prior to the statistical analysis of these data, recorded EEG signals were filtered with a low-pass filter using the fifth level Butterworth filter with a limit of 25Hz frequency. Further, the signal was separated into single units and corrected compared to the baseline in the way that for each unit a subtracted average value was taken from the baseline. Meaning, the average value of the period of 300ms before the onset of each of the stimuli was subtracted from the recorded unit value. Units that contained artefacts produced due to eye movements were excluded from further analysis. Grand-average waveforms for each of the experimental situations were calculated as the average measure of all units of all participants.

Since there was only one cap for all the participants (which due to improper size distorted most of the data from the side electrodes), only data which were collected from F, C and P electrodes were used for the analysis. Prior to statistical analysis, subjects with too many noisy recordings (over 50%) were excluded from the dataset, which left the following sample by phases (Table I-5):

Table I-5: Remaining ERP sample after removal of the cases with too noisy recordings

Phase	Remained sample
I	101
II	94
III	96

Central electrodes that remained in the analysis (C, F and P), were intended for 3x3 factorial design. After removal of additional noisy electrodes, the following distribution of the recorded sample remained (Table I-6):

Table I-6: Number and percentage of remaining sample by electrodes and phases

	Phase I		**Phase II**		**Phase III**		**TOTAL**
	N	%	N	%	n	%	%
C3	92	91,09	87	92,55	89	92,71	92,12
Cz	78	77,23	76	80,85	76	79,17	79,08
C4	95	94,06	90	95,74	93	96,88	95,56
F3	82	81,19	78	82,98	71	73,96	79,38
Fz	49	48,51	51	54,26	43	44,79	49,19
F4	75	74,26	74	78,72	65	67,71	73,56
P3	30	29,70	27	28,72	31	32,29	30,24
Pz	96	95,05	90	95,74	90	93,75	94,85
P4	99	98,02	92	97,87	94	97,92	97,94

Due to an unequal distribution of the remaining electrodes sample, a cut point of 50% validity was implemented (those electrodes with over a 50% sample remained). In this process, two electrodes were removed (P3 and Fz). Since the 3x3 model could not be implemented, it was decided that only the 3 central electrodes should be used for the following two reasons: data from all these three electrodes were preserved and effects of categorisation and language tasks are identifiable at these central electrodes.

For the statistical analysis of the remaining three electrodes (C3, Cz and C4) the difference-waves method was used, meaning that it was calculated the difference between first half and the second half of the exposed sample in the I phase.

For the statistical analysis of the remaining three electrodes (C3, Cz and C4) the difference-waves method was used, meaning that it calculated the difference between the first half and the second half of the exposed sample in phase I. In the second phase, the difference was calculated between the old and new represented stimuli from the same category and in the third phase, the difference between the congruent and incongruent (name and picture) were calculated.

The outliers (over 2SD from both sides) for all three phases (I, II, III) were removed. Variables were close to normal distribution.

For each of the remaining electrodes, a one way ANOVA was implemented for each of the 20ms windows, for each of the phases. Results showed that only significant consecutive windows (more than two) were 16-19 of the electrode C4 (300ms-360ms) in phase I.

Since there were too many comparisons, the Bonfferoni correction was implemented: for each electrode, 49 comparisons were made, which transfers p = .05 into p = .001 (0.05/49=0.001) as a significance level after the correction. However, only one window (16) remained significant after correction.

Since further analysis would not have been necessary after this correction, analysis was conducted without it for illustration purposes. Windows 16 to 19 were averaged for each of the electrode and identified as a P3 component (300ms - 360ms). Average results for each Group for the C4 electrode are presented on the following graph (Figure I-9):

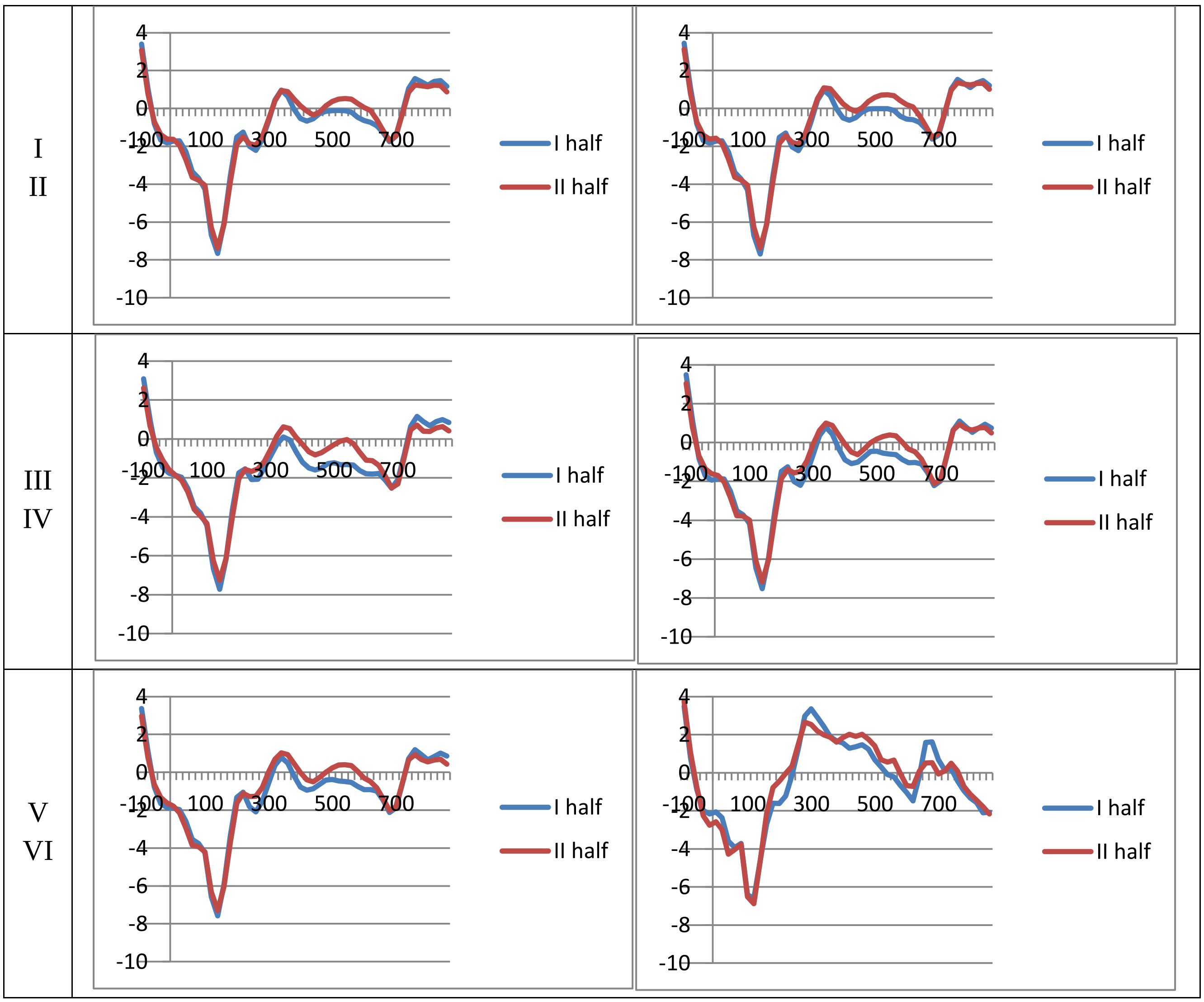

Figure I-9: ERP waves of I phase on C4 electrode divided by groups

Difference waves for C4 are presented in the following graph (Figure I-10):

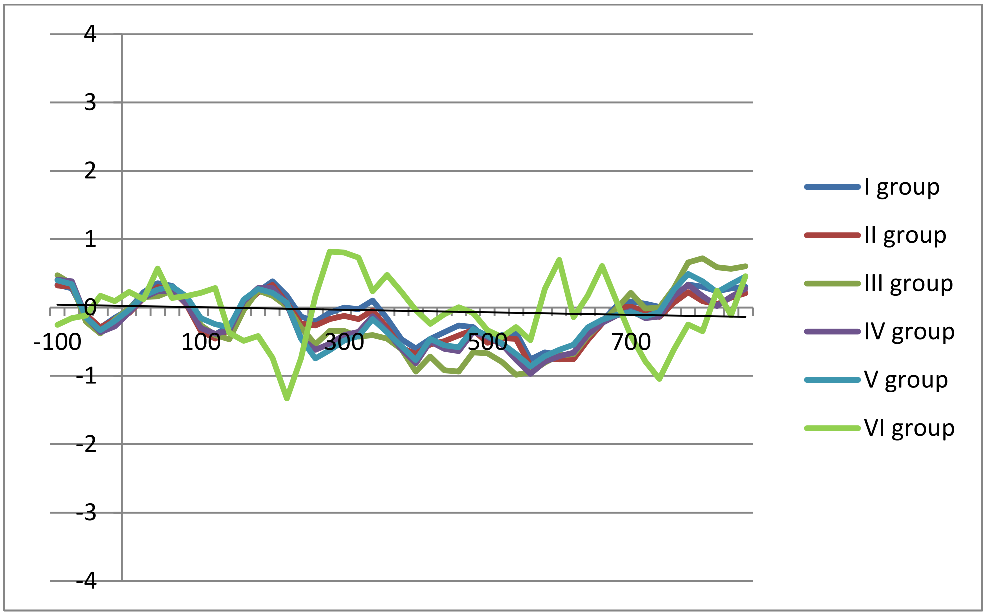

Figure I-10: Difference waves for I phase of C4 electrode

A two-way mixed-design 6x3 ANOVA was conducted with Group as a between-subjects factor (1 to 6) and Electrode (C3, Cz, C4) as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group ($F(5,47) = 1.046$, $p = .402$, $\eta^2 = .100$) and also no significant main effect of Electrode ($F(2,46) = 1.048$, $p = .359$, $\eta^2 = .044$) or significant Group x Electrode interaction ($F(10,94) = 1.271$, $p = .238$, $\eta^2 = .119$).

Absence of results in the first phase could be interpreted in the following way: comparison between the first and the second half of the presented trials were not significantly different, since the main difference was between the first and the rest of blocks. Only in most of the first block trials, participants were guessing responses by chance. This was significantly improved in the second and other blocks. Once values of the first block are averaged with the rest of the first half, this difference diminishes. The difference could possibly be identified between the first and the last ninth block, but we simply did not have enough trials in those blocks for valid comparison.

Regarding the results from the third test phase, we could possibly conclude the following:

1. There are no effects of difference of verbal labels on category learning or any early ERP component. This is not highly probable: many previously stated researches identified at least effects of sound symbolism on ERP components (Kovic, Plunkett & Westermann, 2010; Sučević et al., 2015). More probable explanation of these results is:

2. Timings for ERP recording were not proper. ISI (Inter stimulus interval) between two stimuli in the third phase was 800ms (between the end of the label pronunciation and picture of the stimuli presentation). Since stimuli were in two different modalities (audio and visual), it was expected that longer ISI will provide a proper baseline for ERP recording. However, it seems that this ISI was too long, which resulted in an absence of identification of the experiment effects.

Unfortunately, this improper timing was identified only after all the recordings were completed (for all experiments in this dissertation), since data processing was slower than experiment conduction. Consequently, all further chapters in this dissertation will not contain ERP results.

5. DISCUSSION

Results obtained in *experiment 1* show that participants *learned* categories that were labelled with phonologically maximally different labels significantly *faster* compared to the ones labelled with minimally different and the ones without labels (silent condition). Additionally, the phonologically mid different condition was not significantly different from either the minimal or maximal condition but descriptively was between the two. Based on that we can presume (not claim) that mid different condition lies somewhere in between and consequently leads to faster learning than minimal, and less than maximal. However, to claim this further research would be needed.

As far as *generalization* is concerned (phase II), there is a similar pattern to the learning results, except that level of generalization in the silent condition was different compared to the pattern in the learning condition (interaction). So we can conclude that generalization in the silent condition decreases compared to the other conditions. Finally, in the label test phase, participants in the minimal condition learned labels significantly worse compared to other conditions.

Concerning *experiment 2*, we obtained results which show that there are no differences between phonologically maximally different labels with congruent sound symbolism condition and phonologically maximal different labels without sound symbolism. When these results are compared with the results from the first experiment (maximally different with incongruent sound symbolism condition and minimally different condition), it was obtained that there is a difference between the minimal condition compared to other condition, but not between three maximally different conditions.

In experiment 1, the main hypothesis of the research was confirmed: once categories are labelled with phonologically maximally different labels, they are learned faster and generalized better compared to those labelled with minimally different labels. Furthermore, the generalization of labelled categories is better compared to the non-labelled condition, which goes in line with the hypothesis that labelled concepts are generalized better.

In experiment 2, results showed that effects of label difference could be assigned to the phonological difference, rather than to sound symbolism (which could be absent due to the "ceiling effect").

Interpretation of these results could be taken from two different points: why categories labeled with phonologically more different labels are *learned* better compared to the ones labeled with phonologically more similar labels? Furthermore, why categories labeled with phonologically more different labels are *generalized* better than those labeled with minimally different labels? And finally, as a part of the previous question, why generalization of labeled conditions was better than non-labeled condition?

The first question is related to learning, which could be interpreted through external properties of the categories (but could also include some internal elements). The second question is related to generalization, which could be interpreted through internal conceptual representation.

Learning results could be interpreted with James' proposed hypothesis: effect of labels on category learning is obtained since labels are more different than categories. Maximally different names, adhered to two categories compose entities which are more easily discriminated than the categories solely. Once minimally different labels are adhered to the categories, those entities will not be easily discriminated as those with maximally different labels. Likewise, the former will be learned faster compared to the latter (which was obtained in the experiment), since more different objects are easier discriminated than more similar ones (James, [1931 (1890)]).

Further explanation of results could be obtained from Lupyan's language augmented thought hypothesis. Lupyan claims that linguistic labels are strongly connected with typical features of the concepts. Once the exemplar of the category is activated, it activates the label, which in return activates these typical features on-line. In the process of learning categories and labels, this effect becomes more and more evident in the learning process. There lays the reason why labeled categories are learned faster, but also generalized better than non-labeled ones. In line with this hypothesis goes the result where participants who learned labels better (maximally different of all types) also learned and generalized concepts better. However, Lupyan's theory does not explain why there are effects of label difference levels, which was obtained in this research.

Related to the *generalization* problem and phonological difference of labels, we could again interpret results based on James' hypothesis: once concepts are created and labels learned, they are stored in the long term memory, but somehow interconnected (associated). Activation of one, leads to the activation of the other and vice versa. This activation increases discriminability between concepts, because they do not consist solely on visual representation, but also on the name as an additional feature (labels).

However, this does not mean that representation of concept consists of unified label representation and representation of visible features, since contemporary theories consider that language and thought are separated (Murphy, 2002; Gleitman & Papafragou, 2005; Wolff & Holmes, 2011). Furthermore, most of the present theories accept an exemplar based view of concepts, so there is no one single representation.

Labeled category is represented with concept – exemplars, that are associated with label representation, which mutually activate each other. This view is close to the Lupyan's view, except that learned labels also influence internal discriminability between concepts based on their phonological difference.

In order to **quantify** previous interpretations, we can use logic from Sloutsky's model of Similarity Induction Naming Categorisation in young children – SINC (Sloutsky & Fisher, 2004; Sloutsky & Fisher, 2012). Even though this model was developed for children, it can be used here since it uses multiplicative rule for exemplar differences (Medin, 1975; Medin & Schaffer, 1978) with which object differences can be quantified. Furthermore, it includes quantification of label phonological differences. And finally, even if adults in Sloutksy's experiments (for example Sloutsky, Lo & Fisher, 2001) do rely almost solely on labels in category classification task, in our experiments it is not possible since labels are presented after the participant's response. In a strict empirical sense, this label can be considered as another feature (needed to be learned) which adheres to the conceptual representation and which does not help in solving individual experimental trials.

According to this model, similarity between two exemplars is calculated with the following equation:

$$\text{Sim}\ (i,j) = \lambda^{\beta}\,\upsilon^{B}$$

Where *i* is an item which is compared to item *j*. λ represent attentional weights of a label and υ the same for the visual property (ranging from 0 to 1), while β represent a number of feature mismatches between two labels and B the number of feature mismatches between appearances (visual features) of two items (B).

Another tool that could be used in the description of differences between category exemplars is the previously mentioned Context Model (Medin, 1975), but also Generalized Context Model, proposed by Nosofski (Nosofsky 1992; Nosofsky & Palmeri 1997) and amended by Maddox and Ashby (Ashby & Maddox, 1993).

Both Sloutsky's and the Generalized Context Model - GCM are designed to illustrate a comparison of exemplar to either another exemplar (Sloutsky) or group of exemplars (GCM). For the purpose of explanation of these results, we can establish an alternative view: the model of *Category learning based on the difference level*, which is derived from previous models.

The description of the model starts with a definition of Between Category Difference (BCD). Each category consists of several dimensions – D[3]. Let us assume that these dimensions are values from 0 (no difference) to 1 (maximum difference). In that case, psychological difference between two categories in one dimension we can express with the following equation:

$$diff(\mathrm{I}, \mathrm{J})_{\mathrm{i}} = \mathrm{S_i} * \mathrm{D_i}$$

where I and J are two categories, *i* is a dimension, S is the attentional weight of the dimension *i* (0 – no attention, 1 – maximum attention) and D_i is the difference between categories on the dimension *i*.

We can further specify overall difference of the categories, using multiplicative rule:

$$\mathrm{diff}\ (\mathrm{I}, \mathrm{J}) = \prod_{i=1}^{m} S_i * D_i$$

where ∏ represents product from i=1 to m (number of relevant dimensions). This value will be between 0 (no difference) to 1 (maximum difference).

If we assume that difference between two categories will be learned easier once the overall difference is higher (Nosofsky 1992). In that sense, once this difference level is calculated, probability (meaning ease) that these two categories will be differentiated (learned) can be expressed with the following equation:

$$\mathrm{P}(\mathrm{I}, \mathrm{J}) = \mathrm{diff}\ (\mathrm{I}, \mathrm{J}) = \prod_{i=1}^{m} S_i * D_i$$

where P(I,J) represents probability that categories I and J will be discriminated (learned).

All previous discussion was related to visual features. We can conduct a similar calculation for the Label Difference (LD[4]).

$$diff_L\ (\mathrm{I}_L, J_L) = \prod_{i=1}^{3} S_i * D_i$$

where I_L and J_L are labels of the categories I and J, Si and Di are attentional weights and level of differences on dimension *i*.

[3] Detailed description of between categories difference for each dimension D_i will be specified in discussion section of the dissertation.

[4] Level of difference will be specified in detail in final chapter of the dissertation.

Likewise, the total difference of labeled categories can be specified with the following equation:

$$\mathrm{diff}\ (I,J)_T = \mathrm{diff}(I,J)*\mathrm{diff}(I_L,J_L)$$

and probability of learning labeled category of it as:

$$P\ (I,J) = \mathrm{diff}(I,J)_T = \mathrm{diff}(I,J)*\mathrm{diff}(I_L,J_L)$$

At this point, we still cannot claim whether the space of label difference is only tridimensional and is it orthogonal or not. Further, this model does not specify whether influence of labels on concept similarity is different than influence of visual attributes, meaning whether words have special status compared to visual features.

This attempt of quantification was conducted for the learning process. As far as generalization is concerned, we can assume that a similar representative process is involved: mental representations of individual category exemplars and names are stored and more easily discriminated if they are more different. This would explain why generalization was better for a maximal different condition compared to a minimal different condition: the former were more different than the latter.

Finally, it is necessary to note that this is a kind of speculative quantifying (how things could work). For further confirmation, additional empirical testing with a different research objective would be necessary.

As a final constraint, we can specify that these results are based on presupposition that participants do not have any previous knowledge about categories and labels (which was achieved by use of novel stimuli and labels). In real life, previous knowledge is always used and interferes with the learning of novel categories (Murphy & Medin, 1985).

CHAPTER II

EFFECTS OF NON-VERBAL LABELS ON CATEGORY LEARNING

1. INTRODUCTION

In the introductory chapter, Lupyan's definition of labels was used, as something which is: "a. consistently correlated with a category and b. used to refer to a category" (Lupyan et al., 2007). Meaning, labels could be any entities, no matter if they were verbal, non-verbal or spatial.

While much of the attention was given to verbal labels, there are significantly fewer studies related to non-verbal labels. Reasons for such disproportion could be because verbal labels are more attractive for studying or that studies of non-verbal labels were not always statistically significant, so they are rarely reported in published papers. The second problem is partially correct, since much of the data with non-verbal labels are generated from research that primarily studied verbal labels, but used non-verbal labels as a control of their effects.

Some of these researches were conducted by Gary Lupyan who primarily studied effects of verbal labels, but used non-verbal labels to illustrate superiority of verbal labels. He used spatial non-verbal labels (Lupyan et. al., 2007) and non-verbal sounds (Lupyan & Thompson-Schill, 2012). In both of these researches, superiority of verbal over non-verbal labels were demonstrated, which allowed Lupyan to conclude that verbal labels have a special status.

There is also group of research conducted on children with similar results. For example, Fulkerson and Waxman (Fulkerson & Waxman, 2007) demonstrated that infants between 6 and 12 months are sensitive to verbal, but not non-verbal sounds and that verbal labels are inherently used as category markers. The problem with studies including children is that we can never sure be if the obtained results are valid for adults too.

On the other hand, from the logical point of view and from the previously stated James' hypothesis, there is no reason to believe that verbal labels have some different status from non-verbal labels. What we described as James' hypothesis states that the elements adhered to the concepts could be some other things, like contexts (like in the example with wine) or some other features that could have the same property as a verbal label.

Additionally, concerning the results from the previous chapter, we can see that only highly differentiable labels lead to category learning effects. There is a possibility that research with non-verbal labels were conducted with labels which were not easily discriminable.

Furthermore, the question is whether there are modality specific effects of non-verbal labels? Some researchers suggest that there is some superiority of auditory modality in attention weight, which overshadows visual input (Sloutsky & Fisher, 2012).

In the line with the aims stated in the introductory chapter, the aim of this chapter is to identify effects of level of differences of non-verbal labels on category learning, using behavioural and neuro-physiological measures[5]. Furthermore, these effects were compared to the non-label condition in order to specify effects of non-verbal labels compared to the baseline.

[5] For the reason stated in the previous chapter, these measures are suspended. Since these measures are recorded, using EEG cap, it is important to report it here (along with the hypothesis), since participants (and the examiner) in the experiments were in different position than without cap and recording.

Even though some research showed that there is supremacy of verbal compared to non-verbal labels, in predictions of outcomes we will stick to basic hypothesis used in this dissertation: James' hypothesis. This hypothesis does not differ between verbal and non-verbal labels, so we expect that categories named with maximally different non-verbal labels will be learned faster and generalized better compared to those labelled with minimally different non-verbal labels. Furthermore, we expect that there will be an overall effect of maximally different non-verbal labels, compared to the baseline condition (silent condition).

For this purpose, two experiments with non-verbal labels were designed. In the first, the effects of visual non-verbal label differences were tested. These non-verbal labels were emblems that were described as emblems of the alien's tribes. In the second experiment, the effects of the auditory non-verbal label differences were tested. These auditory labels were digitally modified sounds described as the sound produced by the specific group of aliens. For both experiments, data for silent condition were taken from Chapter I (first 15 participants).

Like in the previous chapter, electro-physiological measures were included (event-related potentials) which are used to measure brain responses in each of the experiment phases (learning, generalization and testing). For the same reasons as in the previous chapter, it is expected that the cognitive load in the first half of the learning phase will be higher compared to the second half, manifested in higher P300 amplitude (Luck, 2014; Polich & Kok, 1995). Furthermore, it is expected that in the label test phase, there will be no difference in semantic expectances between minimal and maximal conditions, which will be manifested in the similar amplitude of N400 component (Luck, 2014; Kutas & Federmeier, 2011).

2. EXPERIMENT 1

In this experiment, the effects of minimally and maximally different visual non-verbal labels on category learning were tested.

Method

Participants

Participants were thirty psychology students, who participated in the experiment for the course credit. Participants were randomly divided into two groups. In the first group, participants were presented with minimally, while in the second group with maximally different visual non-verbal labels. Data for the third group were taken from Chapter I (silent condition).

Material and stimuli

Material and stimuli were the same as those in Chapter I, except labels which were non-verbal (visual in this experiment).

Visual non-verbal labels were taken from Hedge & Van Esen (Figure II-1), with permission (Hedge & Van Esen, 2003; Hedge & Van Esen, 2004). These labels were generated in Matlab 2010b, by the script provided by the authors. Minimally different labels represent textures created by the parameters in a polar coordinate system (marked in this paper by number 42 and 44). One of the maximally different labels was also created in a polar coordinate system (marked as number 44), while the second one was produced in a hyperbolic coordinate system (marked as number 13). Both pairs of labels were presented in Figure II-2.

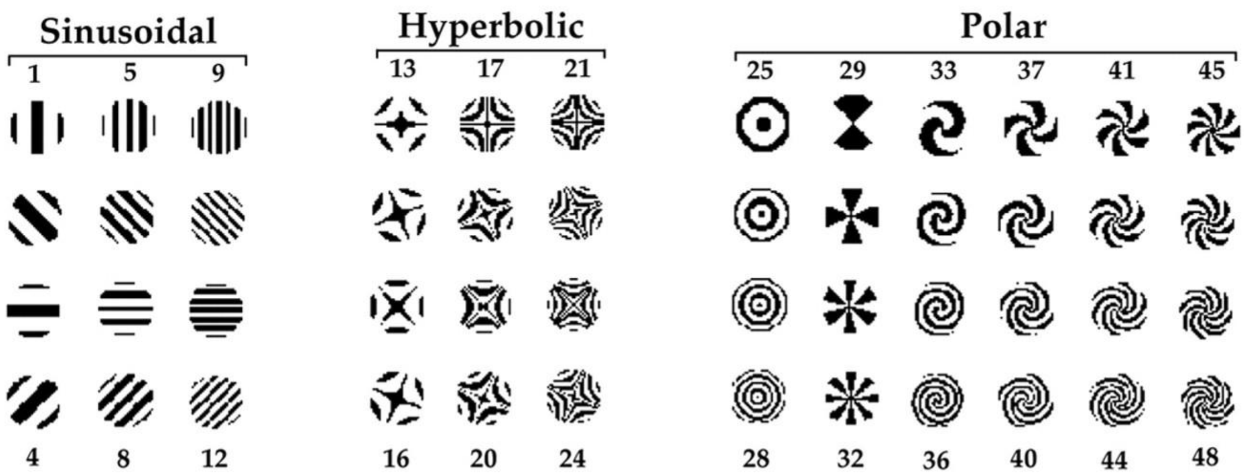

Figure II-1: List of visual stimuli (from Hedge & Van Esen, 2004)

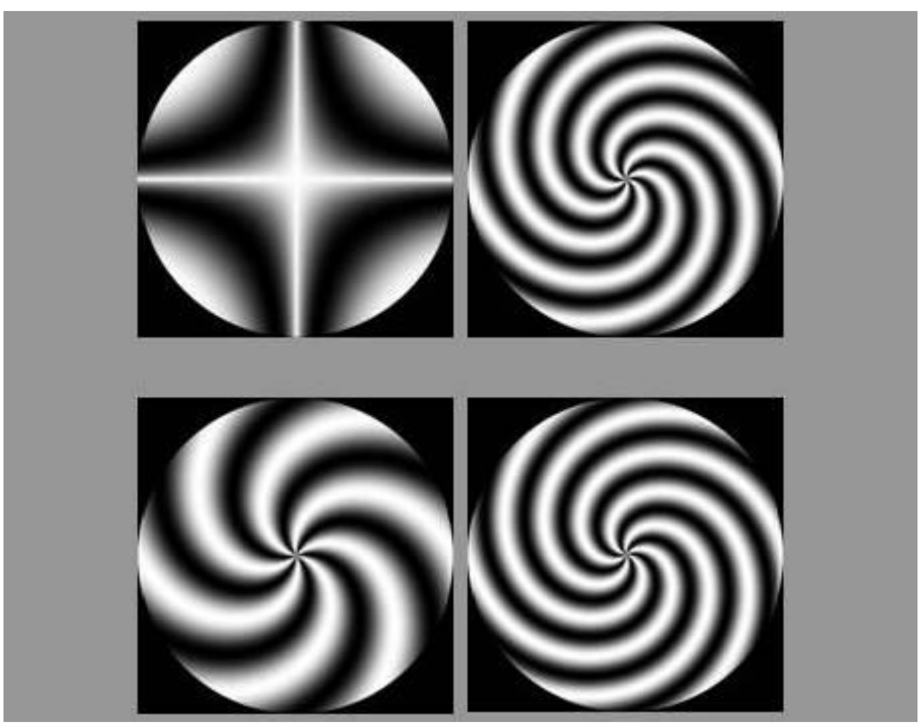

Figure II-2: Visual maximally (upper two) and minimally different labels (lower two)

Design and procedure

The design and procedure were the same as in the previous chapter, except that instead of the verbal label, the participant saw a visual non-verbal label (as shown in the Figure II-3).

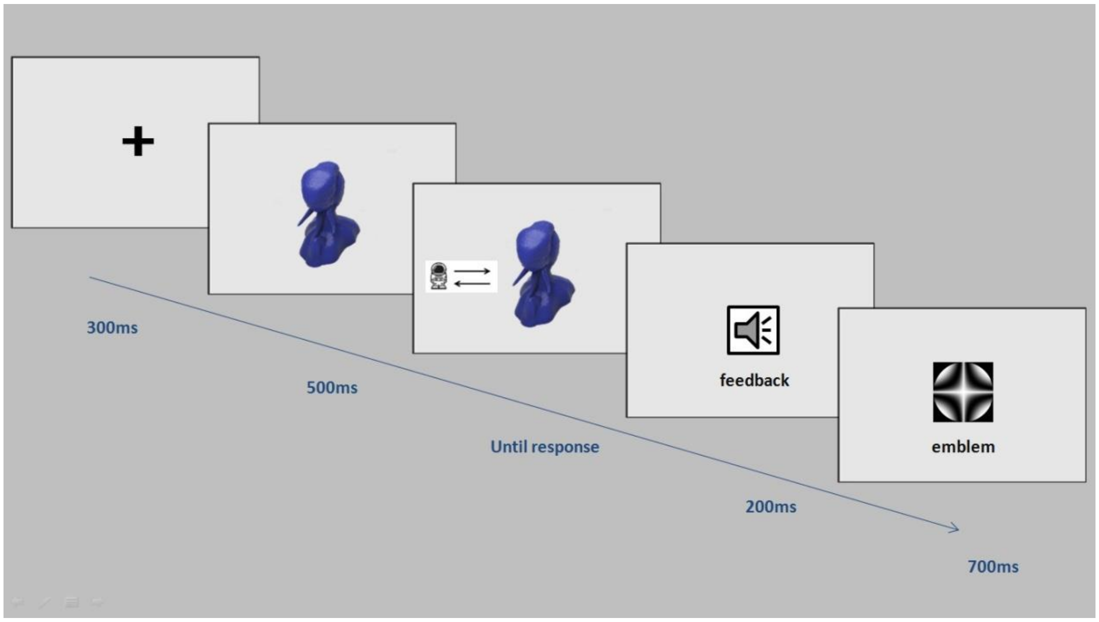

Figure II-3: Trial structure in the learning phase (phase I)

Participants were instructed that the visual non-verbal labels were emblems of the group of aliens (one for good and one for bad). Like in the first chapter, these labels were completely task redundant, since participants could do the task even if they completely ignored them. Again, participants were instructed to pay special attention to these emblems, since their knowledge of it would be tested at the end of the experiment.

In each phase of this experiment, brain potentials were recorded from the onset of the picture of the alien on the screen.

As a dependent variable, the percentage of correct responses (accuracy) was measured. Furthermore, as a control variable, reaction time was recorded. Finally for the ERP measure, measure of voltage was used, expressed in microvolts.

Results and discussion

Behavioural results

Descriptive results in *Reaction time analysis* for each of the phases are presented in the following table (Table II-1):

Table II-1: Descriptive results of reaction times (in milliseconds) in Experiment 1

Group		**N**	**Mean**	**St.Deviation**
Training	nl_VIZ_MIN	15	930.26	387.01
	nl_VIZ_MAX	15	1083.43	427.54
	Silent	15	955.64	307.42
Test	nl_VIZ_MIN	15	980.54	347.78
	nl_VIZ_MAX	15	1034.81	411.08
	Silent	15	790.99	273.60
Label_test	nl_VIZ_MIN	15	804.60	197.07
	nl_VIZ_MAX	15	813.43	334.45

A two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with three levels (minimal, maximal and silent) and Phase as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group ($F(2,42) = 1.111$, $p = .339$, $\eta^2 = .050$) and also no significant main effect of Phase ($F(1,42) = 2.231$, $p = .143$, $\eta^2 = .050$). Additionally, there was no significant Phase x Group interaction ($F(2,42) = 2.916$, $p = .065$, $\eta^2 = .122$).

Finally, there were no significant differences in reaction times in the label test phase ($F(1,28) = .008$, $p = .930$, $\eta^2 = .000$).

These results show that there are no differences in reaction times across conditions and that there was no eventual speed-accuracy trade off.

Accuracy. Descriptive results for accuracy (percentage of correct responses) are presented in the following table and chart (Table II-2 and Figure II-4):

Table II-2: Percentage of correct responses in Experiment 1

Group		**N**	**Mean**	**St.Deviation**
Training	nl_VIZ_MIN	15	68.98	12.27
	nl_VIZ_MAX	15	74.12	11.95
	Silent	15	76.94	9.88
Test	nl_VIZ_MIN	15	75.42	12.69
	nl_VIZ_MAX	15	79.17	11.11
	Silent	15	77.71	15.61
Label_test	nl_VIZ_MIN	15	55.28	9.93
	nl_VIZ_MAX	15	57.38	14.15

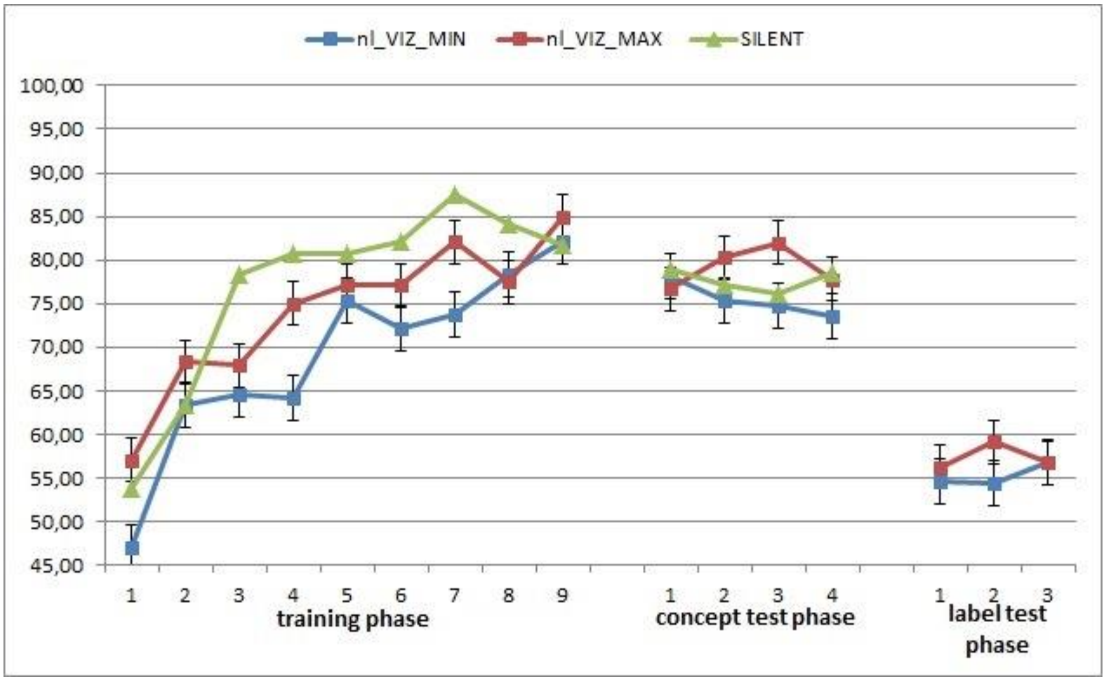

Figure II-4: Percentage of correct responses over phases and blocks in Experiment 1

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between subject factor and Phase as a within-subjects factor (training and test). There was no significant main effect of Group ($F(2,42) = 0.856$, $p = .432$, $\eta^2 = .039$). However, there was a significant main effect of Phase ($F(1,42) = 10.822$, $p = .002$, $\eta^2 = .205$). Additionally, there was no significant Group x Phase interaction ($F(2,42) = 1.892$, $p = .163$, $\eta^2 = .083$).

For the label test phase, results follow the same pattern: there were no significant differences between groups ($F(1,28) = .222$, $p = .641$, $\eta^2 = .008$).

Within subject differences are due to the learning. Similarly to most of the experiments of this type, it is expected that participants will perform better in the test phase when they already know concepts compared to the entire training phase.

These results do not support James' hypothesis. It seems that visual non-verbal labels, no matter if they are minimally or maximally different, do not affect concept learning and generalization. The reason for these effects we can interpret through low level learning of the labels. Additionally, there is possibility that non-verbal labels do not draw attention at the same scale as verbal labels.

3. EXPERIMENT 2

As it was stated in the introduction, in this experiment the effects of auditory non-verbal labels and their effect on category learning and generalization were tested.

Unlike previous labels, for which we used some linguistic or mathematical measures to create minimal and maximal difference, for auditory non-verbal labels, a norming study was conducted in order to identify minimally and maximally different stimuli.

Norming Study

In this study, 8 independent judges participated. Their task was to rate on a 7 point scale (1-completelly same to 7 maximally different) the overall difference between pairs of auditory stimuli they heard. Eight sounds were compared between each-other and the pair with the minimal average score (1.13) was selected as minimally different, while the pair with the maximal average score (6.00) was selected as maximally different.

Method

Participants

The participants were thirty psychology students, who participated in the experiment for the course credit. Participants were randomly divided into two groups. In the first group, participants were presented with minimally different non-verbal auditory labels, while in the second group with maximally different ones. Data for the third group were taken from Chapter I (silent condition).

Material and stimuli

Materials and stimuli were completely the same as in the previous experiment, except that non-verbal auditory labels were used which were maximally and minimally different, based on the results from the norming study.

All stimuli were digitally recorded natural sounds, which were cut and equalized in length (700ms) so that the type of the original sound could not be recognized (for example, the sound of a leaf while walking, which was over two seconds in total was reduced to 700ms). Selected minimally different labels were digitally altered in the programme Sound Forge 7. Digital wave recording of a minimally different pair is presented in the following picture (Figure II-5):

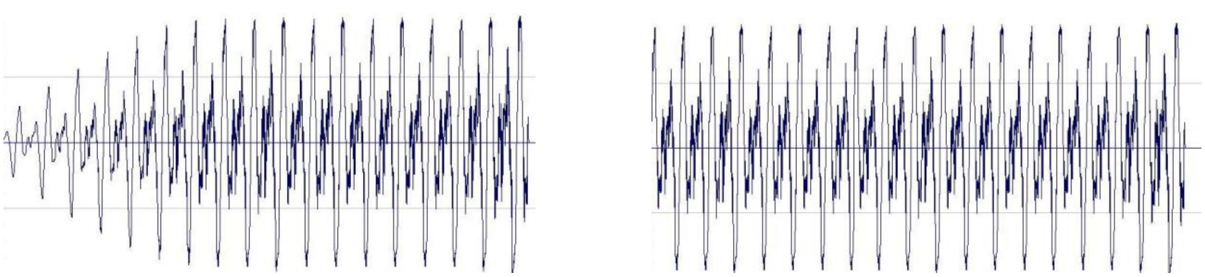

Figure II-5: Wave recordings of minimally different auditory non-verbal labels

In order to make these labels more inherent to the aliens, participants were told that these are sounds that aliens make (like their speech). Meaning, one group of aliens make one type of the sound while the other makes another type of the sound.

Design and procedure

Design and procedure were the same as in the previous experiment.

Results and discussion

Behavioural results

In *Reaction time analysis* for each of the phases the following descriptive results were obtained (Table II-3):

Table II-3: Descriptive results of reaction times (in milliseconds) in Experiment 2

Group		N	Mean	St.Deviation
Training	nl_AUD_MIN	15	1007.27	353.95
	nl_AUD_MAX	15	1142.89	357.33
	Silent	15	955.64	307.42
Test	nl_AUD_MIN	15	965.30	401.63
	nl_AUD_MAX	15	1068.31	439.58
	Silent	15	790.99	273.60
Label_test	nl_AUD_MIN	15	880.65	229.22
	nl_AUD_MAX	15	933.56	400.65

A two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with three levels (minimal, maximal and silent) and Phase as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group ($F(2,42) = 1.958$, $p = .154$, $\eta^2 = .085$), no significant main effect of Phase ($F(1,42) = 3.785$, $p = .058$, $\eta^2 = .083$) and no significant Phase x Group interaction ($F(2,42) = .580$, $p = .564$, $\eta^2 = .027$). Finally, there were no significant differences in reaction times in label test phase ($F(1,28) = .197$, $p = .660$, $\eta^2 = .007$).

These results show that there are no differences in reaction times across conditions and that there was no eventual speed-accuracy trade off.

Accuracy. Descriptive results for accuracy (percentage of correct responses) are presented in the following table and chart (for each phase and within phase blocks):

Table II-4: Percentage of correct responses in Experiment 2

Group		N	Mean	St.Deviation
Training	nl_AUD_MIN	15	76.20	15.12
	nl_AUD_MAX	15	76.16	15.29
	Silent	15	76.94	9.88
Test	nl_AUD_MIN	15	77.43	12.47
	nl_AUD_MAX	15	78.96	18.71
	Silent	15	77.71	15.61
Label_test	nl_AUD_MIN	15	54.72	8.87
	nl_AUD_MAX	15	68.34	17.71

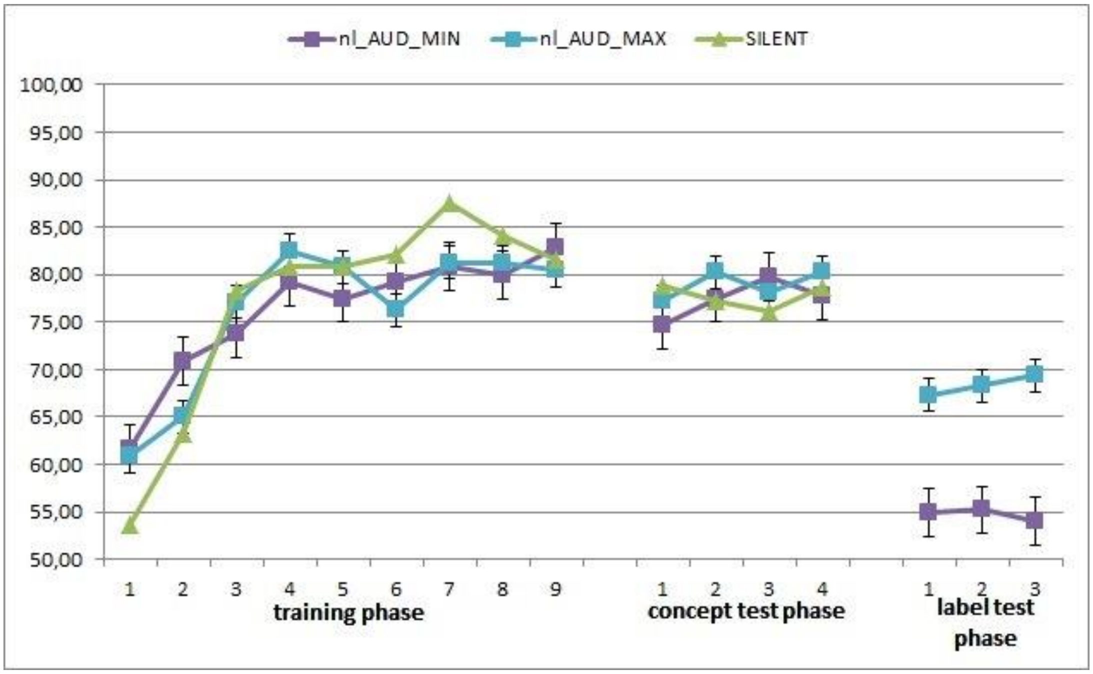

Figure II-6: Percentage of correct responses over phases and blocks in Experiment 2

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between subject factor and Phase as a within-subjects factor (training and test). There was no significant main effect of Group ($F(2,42) = .010$, $p = .990$, $\eta^2 = .000$). Additionally, there was no significant main effect of Phase ($F(1,42) = 2.515$, $p = .120$, $\eta^2 = .057$). Finally, there was no significant Group x Phase interaction ($F(2,42) = .375$, $p = .690$, $\eta^2 = .018$).

For the label test phase, results showed that there is a significant difference between groups ($F(1,28) = 7.081$, $p = .013$, $\eta^2 = .202$) in favour of maximally different non-verbal labels which participants learned much better.

Results showed, like in the previous experiment, that there are no effects of auditory non-verbal labels on category learning. Which is more interesting, there is a significant difference between levels of label learning in the two groups, but this did not affect concept learning and generalization.

4. DISCUSSION

Results obtained in experiment 1 showed that there are no differences between maximally and minimally different visual non-verbal labels on category learning and generalization. Additionally, there are no differences in the level of labels learning between conditions, even though the level of learning of these labels was low (slightly above chance). Finally, none of these labels facilitated category learning or generalization, since there were no differences compared to the silent condition.

In the experiment 2, similar results were obtained: there were no differences between participants who learned categories labelled with minimally different labels compared to those who learned them with maximally different labels. The same pattern was seen with generalization. There is a difference in label learning, since labels were better learned in a maximally different condition. However, this difference did not contribute to better category learning or generalization in this condition. And finally, as in the previous experiment, these labels did not facilitate category learning or generalization, since there were no differences compared to the silent condition.

We can conclude that non-verbal labels do not facilitate category learning or generalization. This does not support James' hypothesis. The question is why are there no effects of non-verbal labels?

As far as *learning* is concerned, first of the possible options is that attentional weights of non-verbal stimuli are not high enough to contribute to better learning. There are no researches which support this claim, but Sloutsky claimed that auditory stimuli overshadow visual ones and leads to higher levels of attention (Sloutsky, 2012). This higher attention levels in return leads to better learning. However, this is not the case here, since we had both auditory and visual non-verbal stimuli, but the effects were the same, except that the auditory maximally different labels were learned better.

This brings us to another possibility: the level of labels learning was low and that brought about minor effects in learning (and generalization). This claim cannot be accepted, since maximally different auditory labels were learned better, but this did not cause better learning (or generalization).

Finally, the most probable possibility is that these labels represented one additional dimension which did influence the overall difference, but not sufficiently enough to be recorded. Based on the equations from the previous chapter, we can represent the effects of non-verbal labels with the following equation:

$$diff_L\,(I_L, J_L) = \prod_{i=1}^{1} S_i * D_i = S_i * D_i$$

As we can see, there is one dimension of non-verbal labels, so the overall difference between two categories, and likewise probability that they will be learned, can be represented as follows:

$$P\,(I,J) = diff(I,J)_T = diff(I,J)*diff(I_L,J_L)$$

Similar logic is used for *generalization*. Once labels and exemplars are learned, the overall difference of their representations will not be that high, so it will not bring any significant difference in generalization.

The key problem in this segment is why there are effects of verbal labels, but there are no effects of non-verbal labels. This is one of the key questions of this dissertation which will be elaborated in the following chapter.

CHAPTER III

RELATION BETWEEN EFFECTS OF VERBAL AND NON-VERBAL LABELS ON CATEGORY LEARNING

1. INTRODUCTION

In the previous two chapters, the effects of the auditory verbal and both auditory and visual non-verbal labels were examined. The question about their relation, or better to say, relation between their effects on category learning is relevant for several theoretical reasons/questions:

- Is there supremacy of certain types of labels and why? This question is important since some authors claim that there is a special status of verbal labels, compared to other types of labels (Lupyan, 2012a). Furthermore, some authors claim that this supremacy of verbal labels compared to non-verbal labels is somehow innate, since labels serve as concept markers (Waxman & Markov, 1995). If it is shown that verbal labels are superior to non-verbal labels, it would go in favour of these hypotheses (but also, it would not prove it). On the other hand, if labels represent just another additional feature (Sloutsky & Fisher, 2004), it would be expected that there would be no difference between verbal or any other (including non-verbal) feature. In this case, we would expect that verbal and non-verbal labels would have the same effects on category learning. However, if there is a different effect of two types of labels, from Sloutsky's perspective, it could be assigned to different attentional weights. This raises another problem related to label modality.

- Is there supremacy of label modality on category learning? Verbal labels are naturally auditory, while visual features are related to visual modality. If we accept the position that the conceptual level is not modality specific, for it is propositional (Pylyshyn, 1973; Fodor & Pylyshyn, 1988) or more probably distributed in nature (McClelland & Rogers, 2003; Rogers & McClelland, 2004), eventual differences across modalities we could expect not to be semantic, but most probably attentional in nature. In Sloutsky's terms, there should be a difference in attentional weights which are related to specific modality. These effects would be recorded on one modality only, independently from the nature of labels (verbal or non-verbal).

Sloutsky's work falls partially in favour of this view, where he interpreted effects of verbal labels as cues which increase attention and consequently better retention of the cues and concepts. The problem with this is that these results are identified only in children, but not in adults (Sloutsky & Fisher, 2012).

- Are labels unmotivated cues? If labels are unmotivated cues, as Edmiston and Lupyan claim (Edmiston & Lupyan, 2015), than any other unmotivated cue could produce the same effect as a verbal label. In our case, non-verbal labels are not in any particular way correlated to any variations within categories which the participants learned, so we can consider them as unmotivated cues. For this reason, potential effects of non-verbal labels, would confirm this hypothesis.

Prior to the discussion of different kinds of label effects, it is necessary to complete an entire set of different labels. The only type of labels that were not examined so far, are written verbal labels. Though, the effects of these types of labels will be examined.

From the point of view of James' hypothesis, the label effects on category learning and generalization are expected. Furthermore, Lupyan used written labels in his research (Lupyan et. al., 2007) and obtained those effects compared to no label condition (silent condition). But what we want to test here is whether there are effects of maximally or minimally different written verbal differences on category learning and generalization. The aims and logic of the study is the same as in the first two chapters.

Like in both previous chapters, electro-physiological measures were included (event-related potentials) which are used to measure brain responses in each of the experiment phases (learning, generalization and testing). For the same reasons as in the previous chapters, it is expected that the cognitive load in the first half of the learning phase will be higher compared to the second half, manifested in higher P300 amplitude (Luck, 2014; Polich & Kok, 1995). Furthermore, it is expected that in the label test phase, there will be no difference in semantic expectances between minimal and maximal conditions, which will be manifested in the similar amplitude of N400 component (Luck, 2014; Kutas & Federmeier, 2011).

2. EXPERIMENT

In this experiment, the effects of phonologically minimally and maximally different verbal written labels on category learning were tested. Additionally, the effects of these labels compared to baseline (silent condition) were tested. Data for the silent condition were taken from Chapter I (first 15 participants).

Method

Participants

Participants were thirty psychology students, who participated in the experiment for the course credit. The participants were randomly divided into two groups. In the first group, participants were presented with minimally, while in the second group with maximally different written verbal labels. Data for the third group were taken from Chapter I (silent condition).

Material and stimuli

Material and stimuli were the same as those in Chapter I, except that labels were written, instead of being spoken (džoset (/dʒoset/) and đoset (/dʑoset/) for minimal difference and ketsi (/ketsi/) and ubom (/ubom/) for maximal difference).

Design and procedure

The design and procedure were the same as in the first chapter, except that instead of the auditory verbal label, participants saw visual verbal labels (as shown in the Figure III-1).

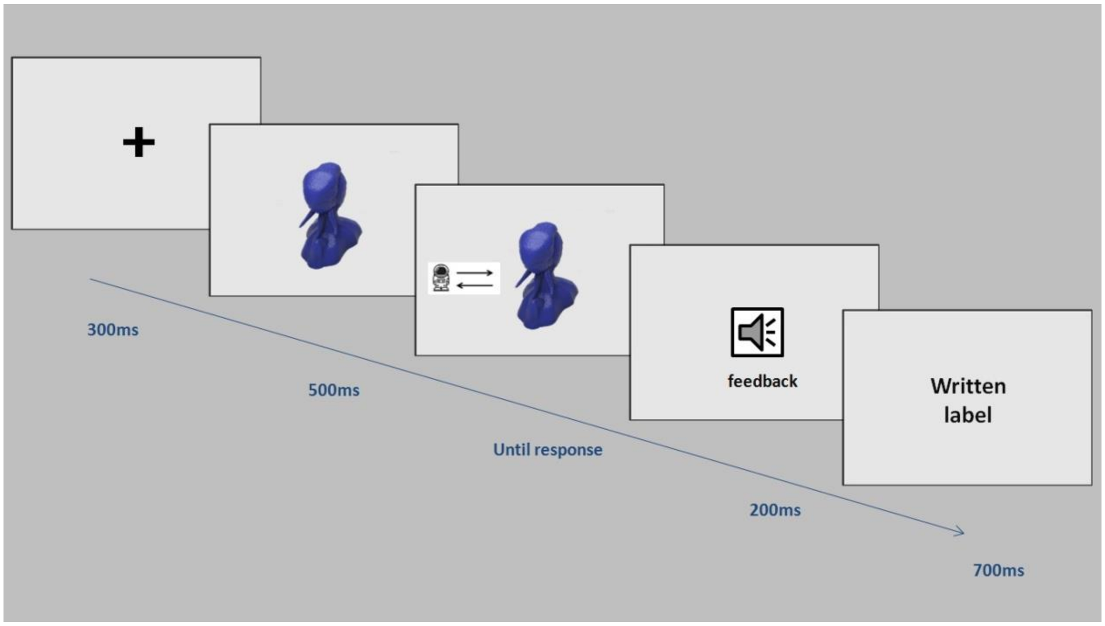

Figure III-1: Trial structure in the learning phase (phase I)

As in the previous experiments, in each phase, brain potentials were recorded from the onset of the picture of the alien on the screen.

Results and discussion

Behavioural results

Descriptive results in *Reaction time analysis* for each of the phases are presented in the following table (Table III-1):

Table III-1: Descriptive results of reaction times (in milliseconds)

Group		N	Mean	St.Deviation
Training	ver_VIZ_MIN	15	914.85	311.53
	ver_VIZ_MAX	15	873.14	191.76
	Silent	15	955.64	307.42
Test	ver_VIZ_MIN	15	881.07	320.08
	ver_VIZ_MAX	15	855.07	251.71
	Silent	15	790.99	273.60
Label_test	ver_VIZ_MIN	15	791.56	227.27
	ver_VIZ_MAX	15	699.78	154.97

A two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with three levels (minimal, maximal and silent) and Phase as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group ($F(2,42) = 0.142$, $p = .868$, $\eta^2 = .007$), but there was a main effect of Phase ($F(1,42) = 4.970$, $p = .031$, $\eta^2 = .106$). Additionally, there was no significant Phase x Group interaction ($F(2,42) = 1.890$, $p = .164$, $\eta^2 = .083$). Finally, there were no significant differences in reaction times in the label test phase ($F(1,28) = 1.670$, $p = .207$, $\eta^2 = .056$).

The differences in Phase are due to learning: in the test phase, since participants learned categories, reacted slightly faster. Between groups, there are no differences in reaction times which mean that there was no eventual speed-accuracy trade off.

Accuracy. Descriptive results for accuracy (percentage of correct responses) are presented in the following table and chart (Table III-2 and Figure III-2):

Table III-2: Percentage of correct responses

Group		N	Mean	St.Deviation
Training	ver_VIZ_MIN	15	67.82	13.50
	ver_VIZ_MAX	15	71.44	10.61
	Silent	15	76.94	9.88
Test	ver_VIZ_MIN	15	73.33	13.16
	ver_VIZ_MAX	15	78.47	11.07
	Silent	15	77.71	15.61
Label_test	ver_VIZ_MIN	15	67.29	15.88
	ver_VIZ_MAX	15	68.47	18.73

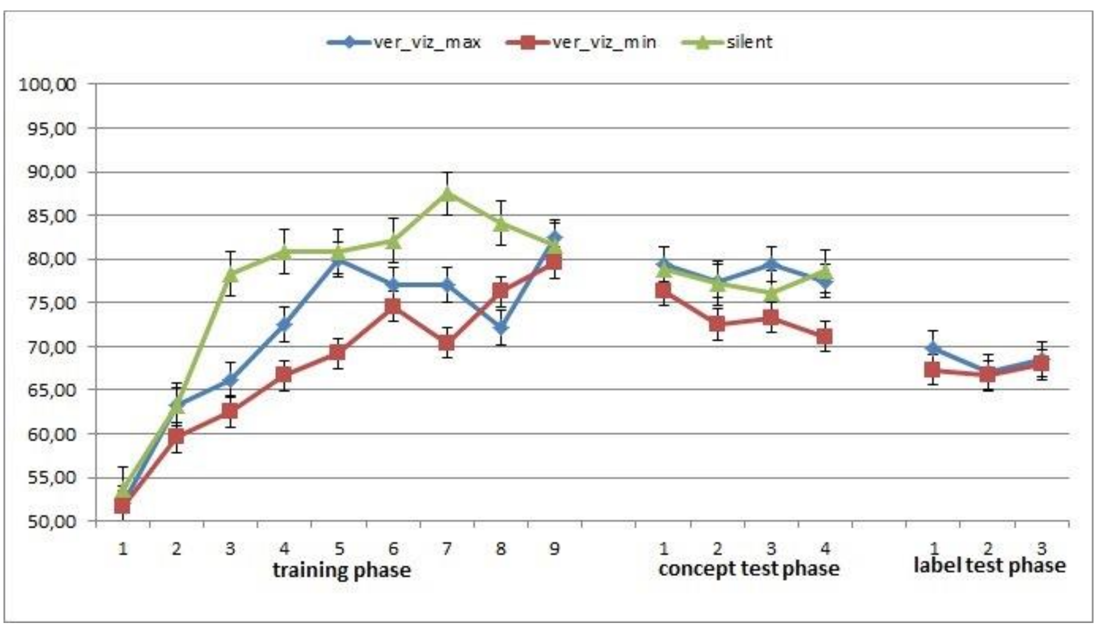

Figure III-2: Percentage of correct responses over phases and blocks

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between subject factor and Phase as a within-subjects factor (training and test). There was no significant main effect of Group ($F(2,42) = 1.317$, $p = .279$, $\eta^2 = .059$). However, there was a significant main effect of Phase ($F(1,42) = 10.099$, $p = .003$, $\eta^2 = .194$). Additionally, there was no significant Group x Phase interaction ($F(2,42) = 1.830$, $p = .173$, $\eta^2 = .080$).

Finally, in the label test phase, there was no significant difference between the two groups ($F(1,28) = 0.035$, $p = .854$, $\eta^2 = .001$).

Within subject differences are due to learning. As in the experiments from previous chapters, it is expected that participants will perform better in the test phase when they already formed concepts compared to the entire training phase.

Surprisingly, these results do not support James' hypothesis. The only difference between this experiment and the first experiment in the first chapter is that we used written, instead of spoken verbal labels. Additionally, Lupyan's results which were obtained in a similarly structured experiment are not replicated.

One of the reasons for this difference could be in the sample size: in Lupyan's experiment, the sample was slightly larger. However, descriptive results obtained in this experiment are not supporting the possibility that a larger sample size could bring to different results.

Another possibility is related to modality specific effects of labels. It might happen that auditory labels have higher attentional weights compared to visual ones. This property made results in the first study more robust.

3. DISCUSSION

The present experiment showed that there are no differences in written verbal labels on category learning. Furthermore, there are no differences between minimally and maximally different conditions.

Lupyan's experiments showed that there are effects of written verbal labels (Lupyan et al., 2007). Furthermore, in the first chapter we showed that there are effects of auditory verbal labels. As we stated, one of the possibilities why this occurred might be due to the small sample compared to Lupyan's experiment. Furthermore, there is a possibility that auditory labels have a higher attentional weights and that they encouraged better learning and generalization.

In the first chapter, we identified effects of auditory verbal labels on category learning and generalization. These effects were not identified with non-verbal labels (chapter II). These effects were absent, even when participants learned some labels better than others. With results from this chapter, all modalities and also verbal and non-verbal labels are examined. Further we can discuss the key question of this chapter: why verbal labels do facilitate category learning and generalization, while non-verbal labels do not? Specifically, what makes words special?

As we noted in the introduction, Edmiston and Lupyan considered verbal labels superior compared to non-verbal labels (Edmiston & Lupyan, 2015). In their experiments, they tested how words or non-verbal sounds, which are strongly associated to category members, activate conceptual knowledge. They showed that the name “dog” does not vary with different exemplars of dogs, while for example, different barks do (bigger dogs bark deeper than small dogs). For this reason, superiority effects of verbal labels are due to this partial association of category members to a non-verbal label (like a bark). Verbal labels are considered to be unmotivated cues, unlike barks which are motivated, since they are associated with one subcategory (those dogs that bark in the presented way).

This interpretation cannot be accepted here, since all category members are associated with a label (visual or auditory, non-verbal or verbal). In these experiments, non-verbal labels are unmotivated cues in all cases, but similar effects as in Lupyan’s experiment were not obtained.

Another interpretation of superiority of verbal labels is given by Waxman and Gelman (Waxman & Gelman, 2009). Unlike non-verbal labels which associate, verbal labels refer to the category. This reference is more abstract, since verbal labels are linked to an abstract representation of an entire category.

From this interpretation, we can conclude that Waxman and Gelman consider concept as a kind of a prototype, or abstract entity to which a name is related. This view is opposite from the one where concept consists of individual exemplars.

There is also an alternative interpretation. If we go back to James’ hypothesis, we can make some amendments on it, with which we will be able to interpret these results. Additional features (like a label) will increase discriminability between two categories. However, it might happen that the incremental effect of this feature will not be that big to bring a significant overall discriminability between categories. On the other hand, labels might increase discriminability in two different ways: by increasing attention or by increasing categories difference on multiple dimensions.

Sloutsky claimed that auditory stimuli can increase attention, since auditory stimuli overshadow visual elements (Sloutsky, 2012). Likewise, in his SINC model, he suggested that attentional weights are higher for auditory stimuli. Since these weights are higher, participants learn better labelled categories (the additional effect of labels is that they bring overall similarity).

In these experiments, we showed that there are no differences between auditory and visual non-verbal labels, so we can consider the argument that auditory stimuli leads to a higher level of attention incorrect. There is possibility that this is the case, but only for the verbal labels, not non-verbal.

On the other hand, there is possibility that verbal labels do not represent only one dimension. It might happen that labels represent multi-dimensional element, which increases dissimilarity between categories named with them. We identified three dimensions of verbal label difference (phonological structure, sonority gradient for alveolar/post-alveolar sounds and vowel position). It might happen that these dimensions combined represent an additional feature for distinction between categories which have a higher influence compared to non-verbal labels.

Concerning the *generalization*, we can follow a similar pattern to learning. Since attention is essential only for learning, we will here focus on labels as a feature of increasing category differences on multiple dimensions. Once stored, non-verbal labels influence discrimination between concepts, but this discrimination is not that high, since there is only one dimension. On the other hand, verbal labels differ on several dimensions, and their influence on overall dissimilarity between the two concepts is higher, which leads to better internal discrimination and also to better generalization. This happens when typical features of the categories are related to several, but not only one dimension. Additionally, each of the three dimensions of verbal labels is associated to typical features of categories, which will additionally emphasise them in the process of activation.

Proposed interpretation could be integrated in the Category based on the difference level model presented in the first chapter. Both category differences and label differences were calculated as in the first chapter, except that label difference was calculated for one dimension only (previous chapter).

$$diff_L\,(\mathrm{I}_L, J_L) = \prod_{i=1}^{1} S_i * D_i = S_i * D_i$$

From the equation it is visible that there is only one dimension, unlike 3 in the equation calculating verbal label difference. This means, if we have similar values for S and D, this difference would be considerably significantly lower compared to the one labelled with verbal labels.

The overall difference would be similarly calculated with the equation from the previous chapter, based on which we can calculate probability that these categories would be learned:

$$\mathrm{P\ (I,J) = diff(I,J)_T = diff(I,J)*diff(I_L,J_L)}$$

However, there is one problem with this kind of interpretation: labels are presented after the feedback was given, not before. It might happen that this feature (label) does not contribute to the difference directly as other visual features do.

If the label was given simultaneously with the object, participants would focus on labels solely and learning would be completed after several trials. In this case, attentional weights of labels would be close to 1 and the overall difference would be high, which would lead to instant learning. On the other hand, even if labels are not presented along with visual stimuli, they do need to be learned and are associated with exemplars in the learning process. In order to describe this influence on the difference, we can assign lower attentional weights to the labels compared to the case when they are presented with a visual object. However, at this moment we can only speculate about the values of these attentional weights.

Another potential problem which could influence the results and conclusions of these experiments is the fact that participants did not learn labels that well, except for the maximal auditory condition. However, the fact that participants learned maximal auditory labels better that the others, did not lead to faster learning or higher generalization, as it was in the case of the experiments with verbal labels. From this we can conclude that non-verbal labels do not have influence on learning or generalization, no matter if they are learned or not.

Further discussion of attentional weights and labels learning in general, will be further specified in the following chapter.

CHAPTER IV

EFFECTS OF ATTENTION ON LABELS INDUCED BY EXPERIMENTAL INSTRUCTION ON CATEGORY LEARNING

1. INTRODUCTION

In previous chapters, it was demonstrated that auditory verbal labels, unlike non-verbal labels, produce effects on category learning and generalization. Those effects were assigned to label learning which consequently leads to a higher conceptual difference. Apart from objective differences of categories, learning also relies on attention. The reason why attention is important is its high correlation with learning. Some classical research on attention, which used the shadowing task (Cherry, 1953; Broadbent, 1958) tested effects of attention through learning. Results showed that a more attended object will be more easily discriminated and learned faster.

In the calculation model presented in previous chapters that described category learning, one of the elements contributing to category difference/ learning was attention or attentional weights. It is supposed that once these weights are higher, participants will be more capable to identify differences between categories with certain features and consequently learn those categories easier.

According to the model, each visual feature and label had its attentional weights. Hence, we can identify two types of attention contributing to the learning: *attention to visual properties* and *attention to labels*.

As far as visual properties are concerned, especially those of living creatures (real or imaginary), some parts of the object are more attended to than others (for example, Batinic, Lalic, Taxitari & Kovic, 2015). Meaning, the head is usually more attended to than the tail or legs. Furthermore, this attention depends on individual strategies of the participants. However, since visual properties in all previous experiments were the same across conditions, we can consider that eventual differences in strategies would be averaged and consequently diminished. A more interesting problem is attention to the labels, since in this dissertation, labels are different across experiments.

In previously stated research (Batinic, Lalic, Taxitari & Kovic, 2015) the effects of labels on category learning were tested. Results showed that participants did not learn better categories that were labelled, but those in which certain features were more salient. Similarly to the experiments in this dissertation, labels were completely task redundant. However, participants in this study were not instructed to learn labels, so they could simply ignore them and complete the task independently.

We can presume that attention to labels in the mentioned study was lower than in the experiment presented in this dissertation. One of the questions this research raises is: what are the effects of lower attention to labels in categorization tasks in general? One possibility is that labels will be ignored and categories will be learned in the same way as without labels. In this way, effects of labels identified in previous experiments will not be recorded. If effects are identified, they can be assigned to higher attention to visual features induced by the simple presence of labels.

A further possibility is that labels will not be learned at the same level as when attention is high, which will consequently lead to a lower generalization (as presented in the previous chapters). Finally, there is a possibility that labels will be learned, no matter if participants are instructed to learn them or not. In this way, similar effects as in the previous experiments in this dissertation should be obtained.

From a theoretical perspective, we can predict some of the outcomes of the experiment without instruction. Relying on James' hypothesis, we can presume that once labels are less attended, the effects on learning of labelled categories will be smaller compared to the cases when labels are attended. Consequently, effects of less attended labels on category learning will not be recorded or will be significantly smaller compared to the effects of more attended ones. Additionally, these expectations could be deduced from Lupyan's hypothesis, according to which labels influence concepts only once they are learned (Lupyan, 2012a; Lupyan, 2012b; Lupyan & Thompson-Schill, 2012).

As far as non-verbal labels are concerned, there is a similar view. From James' perspective, we could not expect effects of labels for the same reason as in the previously stated expectation. Furthermore, since results from Chapter II showed that there were no effects of non-verbal labels when experimental instruction was present, we also do not expect these effects to be obtained once instruction is not present and consequently labels are not attended.

The aim of this chapter is to identify effects of absence of experimental instruction (meaning lower attention to labels) on category learning. It is expected that the effect of labels on category learning and generalization will be identified only if there are explicit instructions given to the participants to learn the labels. If participants do not get the explicit instruction, the effects should be smaller or non-existent.

For this purpose, three experiments were designed in which the effects of instruction on category learning were tested: for verbal labels, spoken and written (experiment 1 and 2) and non-verbal labels (experiment 3). In experiment 3, only visual non-verbal labels were used, since in Chapter II, effects of both types of non-verbal labels were not obtained even when instruction was present.

The design and stimuli were completely replicated from the previous chapters, except that participants did not receive instruction to pay attention to and learn labels.

2. EXPERIMENT 1

In this experiment, the effects of minimally and maximally different auditory verbal labels without instruction on category learning were tested.

Method

Participants

Participants were thirty psychology students, who participated in the experiment for the course credit. Participants were randomly divided into two groups. In the first group, participants were presented with minimally, while in the second group with maximally different spoken verbal labels. Data for the no labelled condition were taken from Chapter I (the first 15 participants).

Material and stimuli

Material and stimuli were completely the same as those in chapter one (minimally and maximally different).

Design and procedure

The design and procedure were the same as in the first chapter, except that participant received different instruction. Instead of getting the instruction to pay special attention to labels and to make efforts to remember them (as in all previous experiments), they received no instruction related to labels, except that labels will be presented.

Results and discussion

Descriptive results in *Reaction time analysis* for each of the phases are presented in the following table (Table IV-1):

Table IV-1: Descriptive results of reaction times (in milliseconds) in Experiment 1

Group		N	Mean	St.Deviation
Training	NI_ver_AUD_MIN	15	1064.64	530.80
	NI_ver_AUD_MAX	15	905.47	254.76
	Silent	15	955.64	307.42
Test	NI_ver_AUD_MIN	15	1012.50	346.24
	NI_ver_AUD_MAX	15	910.69	255.58
	Silent	15	790.99	273.60
Label_test	NI_ver_AUD_MIN	15	779.65	221.01
	NI_ver_AUD_MAX	15	922.25	401.60

A two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with three levels (minimal, maximal and silent) and Phase as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group ($F(2,42) = 1.086$, $p = .347$, $\eta^2 = .049$). However, there was a significant main effect of Phase ($F(2,42) = 4.641$, $p = .037$, $\eta^2 = .100$). Additionally, there was no significant Phase x Group interaction ($F(2,42) = 2.323$, $p = .110$, $\eta^2 = .100$).

Finally, there was no significant difference in reaction times in the label test phase ($F(1,28) = 1.452$, $p = .238$, $\eta^2 = .049$).

These results show that there are no differences in reaction times across groups and that there was no eventual speed-accuracy trade off. The differences between phases are due to learning: once participants learn categories, they tend to react much faster than in the training phase.

Accuracy. Descriptive results for accuracy (percentage of correct responses) are presented in the following table and chart (Table IV-2 and Figure IV-1):

Table IV-2: Percentage of correct responses in Experiment 1

Group		N	Mean	St.Deviation
Training	NI_ver_AUD_MIN	15	71.16	13.64
	NI_ver_AUD_MAX	15	76.90	13.72
	Silent	15	76.94	9.88
Test	NI_ver_AUD_MIN	15	78.47	16.31
	NI_ver_AUD_MAX	15	76.60	18.43
	Silent	15	77.71	15.61
Label_test	NI_ver_AUD_MIN	15	70.35	18.31
	NI_ver_AUD_MAX	15	73.23	15.52

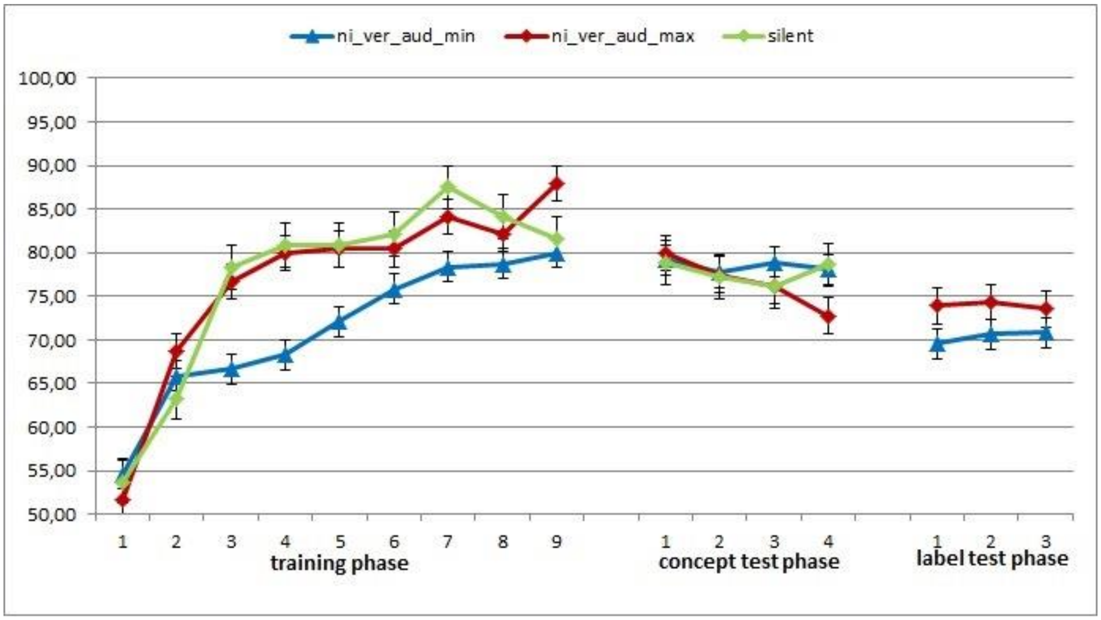

Figure IV-1: Percentage of correct responses over phases and blocks in Experiment 1

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor and Phase as a within-subjects factor with two levels (training and test).

There was no significant main effect of Group ($F(2,42) = .131$, $p = .878$, $\eta^2 = .006$). Furthermore, there was a marginally significant main effect of Phase ($F(1,42) = 3.426$, $p = .071$, $\eta^2 = .075$). Finally, there was a marginally significant Group x Phase interaction ($F(2,42) = 2.889$, $p = .067$, $\eta^2 = .121$, shown in the Figure IV-2). This marginally different interaction is due to the relation between maximally and minimally different conditions. This is visible when the silent condition is removed – interaction reaches significance ($F(1,28)=4.453$, $p = .044$, $\eta^2=.137$).

In the label test phase, there was no significant difference between the two groups ($F(1,28) = .216$, $p = .645$, $\eta^2 = .008$).

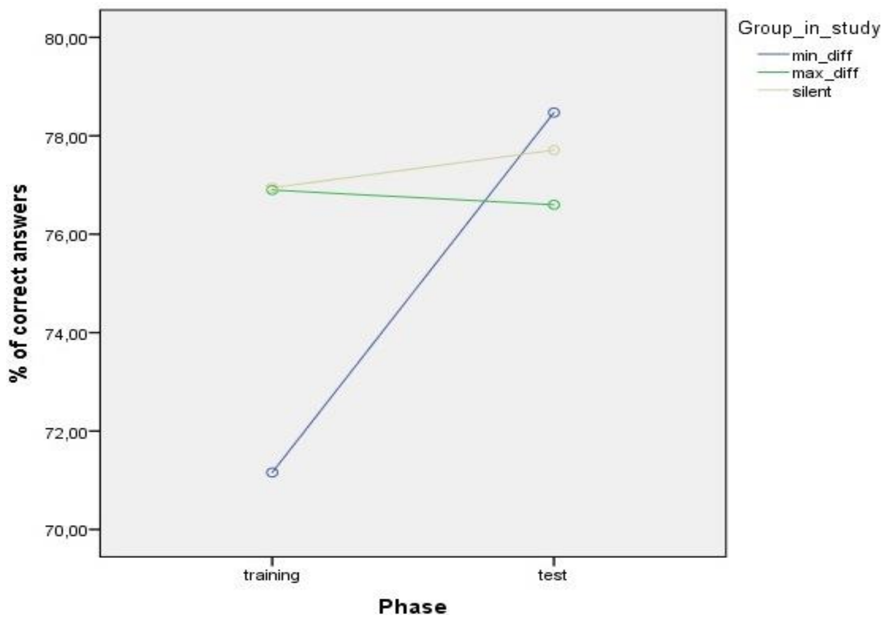

Figure IV-2: Interaction between training and test phase in Experiment 1

The obtained results showed that the effects on category learning we found in Chapter I disappear once instruction is removed. Additionally, there is a marginally significant interaction between Phase and Group. This interaction (as we can see from the Figure IV-2) is due to the minimal difference condition, since there is a lower score in the training compared to test phase, while the other two conditions have a similar pattern for both phases. Meaning, once there is no instruction, participants learn poorer if labels are minimally different, but generalization is the same, no matter if participants learned concepts labelled with maximally or minimally different labels (or not labelled). This data is even more interesting if we notice that there is no difference between groups in the level of learned labels, since there are no differences in the label test phase. It seems that similar level of labels learning leads to same generalization, but a minimally different condition brings some kind of an inhibition in the learning process, even though participants were not instructed to pay attention to the category labels.

Comparison with results from the experiment with instruction

In order to get complete picture of the effects of the instruction, we need to compare the results from experiments where the instruction is given with those where instruction is not given. For this purpose, results from this experiment and results from Experiment 1 from the first chapter will be compared. These two experiments are completely the same, except the difference in instructions. The first 15 participants for each condition (minimal and maximal) from the first experiment in the first chapter were taken for the further analysis.

For reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with four levels and Phase as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group ($F(3,56) = 1.654$, $p = .187$, $\eta^2 = .081$). Furthermore, there was no significant main effect of Phase ($F(1,56) = 1.208$, $p = .276$, $\eta^2 = .021$). Additionally, there was no significant Phase x Group interaction ($F(3,56) = 0.861$, $p = .467$, $\eta^2 = .044$). Finally, there was no significant difference in reaction times in the label test phase ($F(3,56) = 2.519$, $p = .067$, $\eta^2 = .119$).

These results showed that participants responded with similar speed across conditions. This means that there was no speed-accuracy trade off.

Accuracy. Descriptive results for all five experimental conditions are together presented in the following graph:

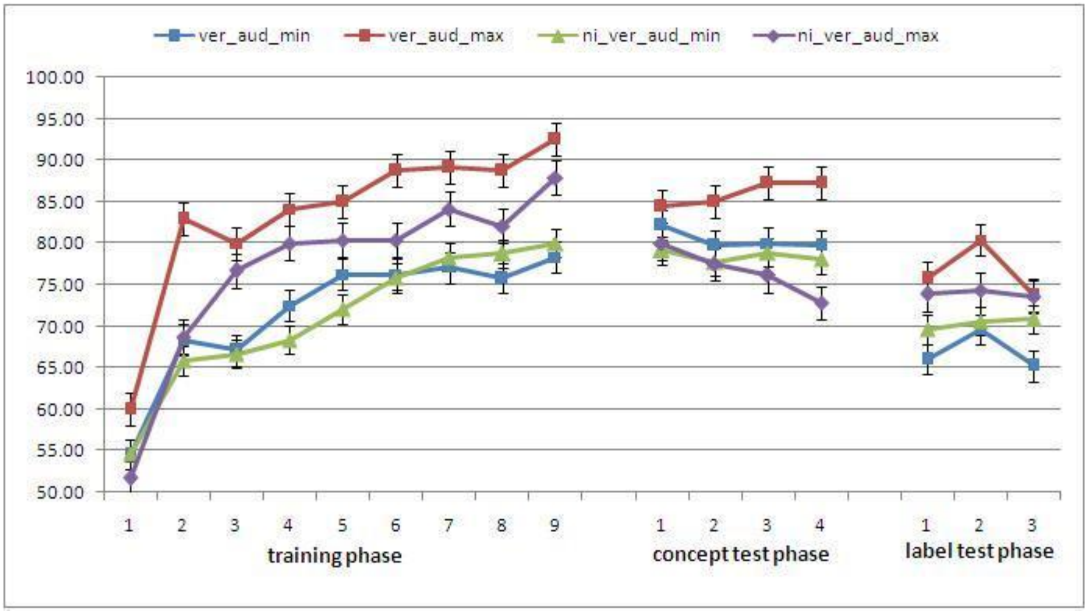

Figure IV-3: Comparison of percentage of correct responses over phases and blocks

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with four levels (minimum and maximum difference in both experiments) and Phase as a within-subjects factor with two levels (training and test).

There was no significant main effect of Group ($F(3,56) = 2.029$, $p = .120$, $\eta^2 = .098$). However, there was a significant main effect of Phase ($F(1,56) = 18.562$, $p < .001$, $\eta^2 = .249$). Finally, there was a significant Group x Phase interaction ($F(3,56) = 3.907$, $p = .013$, $\eta^2 = .173$, shown in the Figure IV-4).

In the label test phase, there was no difference between groups ($F(3,56) = 1.057$, $p = .375$, $\eta^2 = .054$).

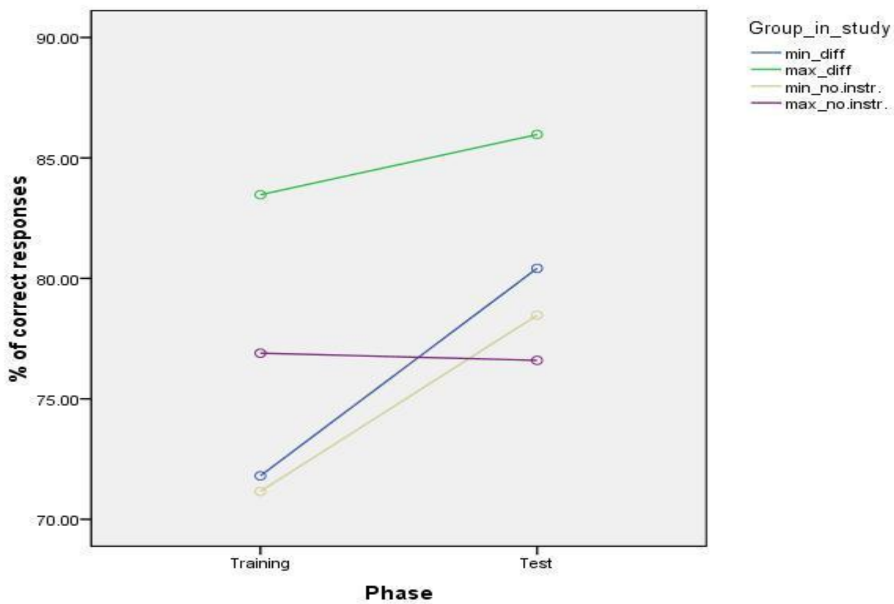

Figure IV-4: Interaction between training and test phase

From the Figure IV-4, it is obvious that interaction is induced by a maximally visual difference, which does not follow similarity between phase patterns as in the rest of the conditions. It seems that maximally different labels without instruction do not bring a higher generalization compared to learning as other conditions. Furthermore, concerning the relation of previously stated effects of minimally different labels without instruction and its relations to the ones with instruction, we can presume that there is possibility that minimally different labels in general lead to category learning inhibition. However, since labels lead to better generalization, this effect is recorded in the test phase. This is obvious with maximally different labels without instruction where generalization is lower since labels are less attended and learned.

3. EXPERIMENT 2

In this experiment, the effects of minimally and maximally different written verbal labels without instruction on category learning were tested.

Method

Participants

The participants were thirty psychology students, who participated in the experiment for the course credit. Participants were randomly divided into two groups. In the first group, participants were presented with minimally, while in the second group with maximally different written verbal labels. Data for the no labelled condition were taken from Chapter I (the first 15 participants).

Material and stimuli

The material and stimuli were completely the same as those in Chapter III.

Design and procedure

The design and procedure were the same as in the previous chapter, except that participants received different instruction. Instead of getting the instruction to pay special attention to labels and to make an effort to remember them (as in all previous experiments), they received no instruction related to labels, except that labels will be presented.

Results and discussion

Descriptive results in *Reaction time analysis* for each of the phases are presented in the following table (Table IV-3):

Table IV-3: Descriptive results of reaction times (in milliseconds) in Experiment 2

Group		**N**	**Mean**	**St.Deviation**
Training	NI_ver_VIZ_MIN	15	1129.90	275.83
	NI_ver_VIZ_MAX	15	979.36	354.99
	Silent	15	955.64	307.42
Test	NI_ver_VIZ_MIN	15	1015.28	311.04
	NI_ver_VIZ_MAX	15	975.65	342.83
	Silent	15	790.99	273.60
Label_test	NI_ver_VIZ_MIN	15	881.30	210.03
	NI_ver_VIZ_MAX	15	848.34	299.43

A two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with three levels (minimal, maximal and silent) and Phase as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group ($F(2,42) = 1.700$, $p = .195$, $\eta^2 = .075$), but there was a significant main effect of Phase ($F(1,42) = 10.052$, $p = .003$, $\eta^2 = .193$). Additionally, there was no significant Phase x Group interaction ($F(2,42) = 2.555$, $p = .090$, $\eta^2 = .108$).

Finally, there was no significant difference in reaction times in the label test phase ($F(1,28) = 0.122$, $p = .730$, $\eta^2 = .004$).

These results show that there are no differences in reaction times across groups and that there was no eventual speed-accuracy trade off. Again, there was a difference between phases, which we can presume is due to learning.

Accuracy. Descriptive results for accuracy (percentage of correct responses) are presented in the following table and chart (Table IV-4 and Figure IV-5):

Table IV-4: Percentage of correct responses in Experiment 2

Group		N	Mean	St.Deviation
Training	NI_ver_VIZ_MIN	15	73.43	9.49
	NI_ver_VIZ_MAX	15	69.12	16.70
	Silent	15	76.94	9.88
Test	NI_ver_VIZ_MIN	15	81.46	10.98
	NI_ver_VIZ_MAX	15	75.35	22.14
	Silent	15	77.71	15.61
Label_test	NI_ver_VIZ_MIN	15	68.41	20.13
	NI_ver_VIZ_MAX	15	69.17	19.34

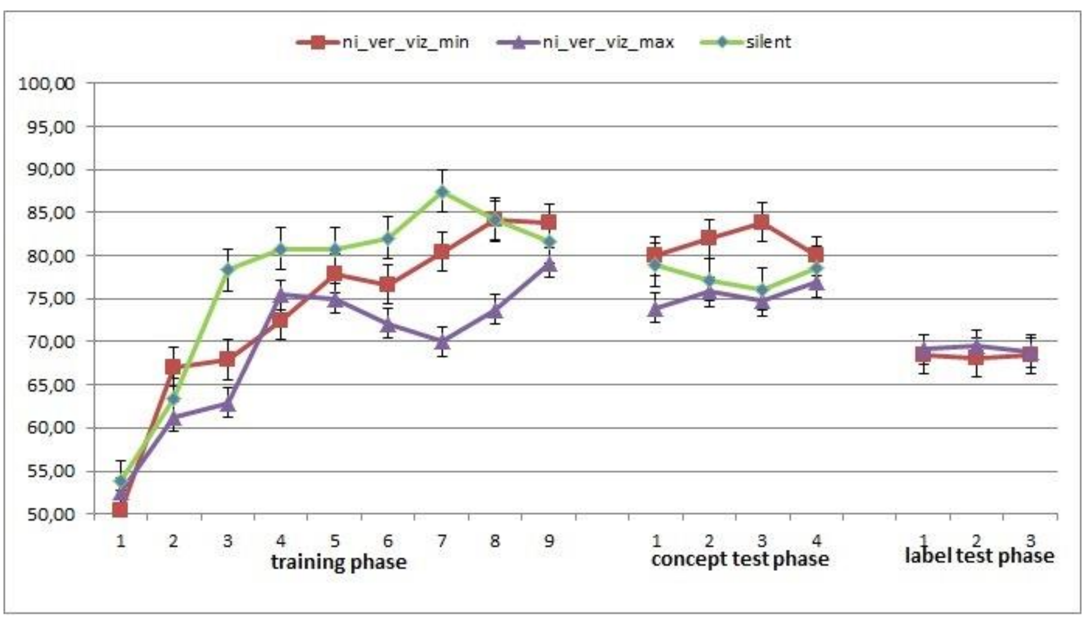

Figure IV-5: Percentage of correct responses over phases and blocks in Experiment 2

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor and Phase as a within-subjects factor with two levels (training and test).

There was no significant main effect of Group ($F(2,42) = 0.660$, $p = .522$, $\eta^2 = .030$). However, there was a significant main effect of Phase ($F(1,42) = 14.705$, $p < .001$, $\eta^2 = .259$). Finally, there was a marginally significant Group x Phase interaction ($F(2,42) = 2.800$, $p = .072$, $\eta^2 = .118$). Unlike with the previous experiment, this marginal interaction is induced with a silent condition, since when this condition is removed, interaction is not significant ($F(1,28) = 0.304$, $p = .586$, $\eta^2 = .011$).

In the label test phase, there was no significant difference between groups ($F(1,28) = .011$, $p = .916$, $\eta^2 = .000$).

These results showed that there is no significant difference between minimal and maximal group, no matter which type of instruction the participants received. This is the same pattern of results as in chapter three, where instructions were presented. To further specify these effects, we need to compare them with results from chapter three.

Comparison with results from the experiment with instruction

Similarly, as in previous experiment, results will be compared to the experiment with instruction from the third chapter.

For *reaction times analysis*, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with four levels (minimum and maximum difference in both experiments) and Phase as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group ($F(3,56) = 1.628$, $p = .193$, $\eta^2 = .080$). Furthermore, there was no significant main effect of Phase ($F(1,56) = 2.675$, $p = .108$, $\eta^2 = .046$). Additionally, there was no significant Phase x Group interaction ($F(3,56) = 0.909$, $p = .443$, $\eta^2 = .046$).

Finally, there were no significant differences in reaction times in the label test phase ($F(3,56) = 1.810$, $p = .156$, $\eta^2 = .088$).

The results showed that participants responded with similar speed across conditions. This means that there was no speed-accuracy trade off.

Accuracy. Descriptive results for all five experimental conditions are presented together in the following plot:

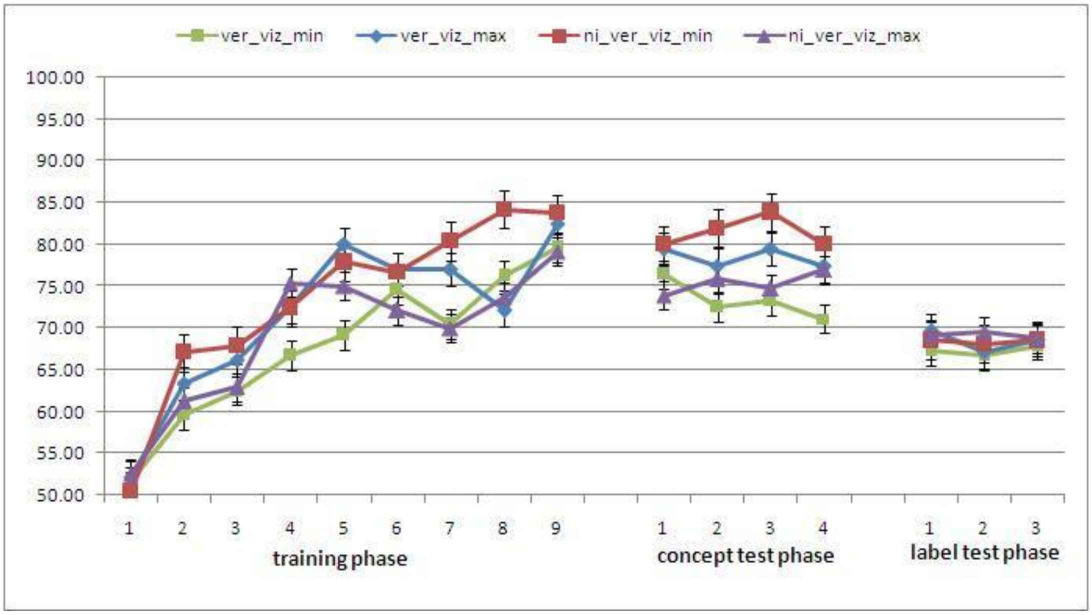

Figure IV-6: Comparison of percentage of correct responses over phases and blocks

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with four levels (minimum and maximum difference in both experiments) and Phase as a within-subjects factor with two levels (training and test).

There was no significant main effect of Group ($F(3,56) = 0.788$, $p = .506$, $\eta^2 = .040$). However, there was a significant main effect of Phase ($F(1,56) = 30.422$, $p < .001$, $\eta^2 = .352$). Finally, there was no significant Group x Phase interaction ($F(3,56) = 0.199$, $p = .896$, $\eta^2 = .011$).

In the label test phase, there was no difference between groups ($F(3,56) = 0.026$, $p = .994$, $\eta^2 = .001$).

It seems that written verbal labels with instructions do not have influence on category learning, since in all four conditions, results have similar pattern. This is not in favour of James' hypothesis and Lupyan's experimental results. It seems that effects of verbal labels are tied to modality: verbal labels can influence concepts learning and generalization only if they are spoken. The possible reason is attention, which is higher only when we have spoken labels.

4. EXPERIMENT 3

In this experiment, the effects of minimally and maximally different visual non-verbal labels on category learning without instruction were tested. Since in chapter two the effects of non-verbal labels were not obtained, only one type of non-verbal labels is tested here: visual non-verbal labels.

Method

Participants

The participants were thirty psychology students, who participated in the experiment for the course credit. Participants were randomly divided into two groups. In the first group, participants were presented with minimally, while in the second group with maximally different visual non-verbal labels. Data for the no labelled condition were taken from Chapter I (the first 15 participants).

Material and stimuli

The material and stimuli were completely the same as those in Experiment 1 in Chapter II.

Design and procedure

The design and procedure were the same as in the chapter two, except that participants received different instruction. Instead of being given the instruction to pay special attention to labels and to make efforts to remember them (as in all previous experiments), they received no instruction related to labels, except that labels will be presented.

Results and discussion

The descriptive results in *Reaction time analysis* for each of the phases are presented in the following table (Table IV-5):

Table IV-5: Descriptive results of reaction times (in milliseconds) in Experiment 3

Group		**N**	**Mean**	**St.Deviation**
Training	NI_nvr_VIZ_MIN	15	967.23	371.81
	NI_nvr_VIZ_MAX	15	989.72	349.90
	Silent	15	955.64	307.42
Test	NI_nvr_VIZ_MIN	15	961.28	340.05
	NI_nvr_VIZ_MAX	15	896.32	331.12
	Silent	15	790.99	273.60
Label_test	NI_nvr_VIZ_MIN	15	803.62	114.31
	NI_nvr_VIZ_MAX	15	888.17	285.15

In reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with three levels (minimal, maximal and silent) and Phase as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group (F(2,42) = 0.346, p = .709, $\eta^2 = .016$), but there was a significant main effect of Phase (F(1,42) = 7.762, p = .008, $\eta^2 = .156$). Additionally, there was no significant Phase x Group interaction (F(2,42) = 2.111, p = .134, $\eta^2 = .091$).

Finally, there were no significant differences in reaction times in the label test phase (F(1,28) = 1.136, p = .296, $\eta^2 = .039$).

Again, these results show that there are no differences in reaction times across groups and that there was no eventual speed-accuracy trade off. The difference in phases is due to learning.

Accuracy. Descriptive results for accuracy (percentage of correct responses) are presented in the following table and chart (Table IV-6 and Figure IV-7):

Table IV-6: Percentage of correct responses in Experiment 3

Group		N	Mean	St.Deviation
Training	NI_nvr_VIZ_MIN	15	73.47	11.30
	NI_nvr_VIZ_MAX	15	69.40	13.05
	Silent	15	76.94	9.88
Test	NI_nvr_VIZ_MIN	15	76.60	17.06
	NI_nvr_VIZ_MAX	15	73.82	13.19
	Silent	15	77.71	15.61
Label_test	NI_nvr_VIZ_MIN	15	61.11	18.64
	NI_nvr_VIZ_MAX	15	53.06	17.13

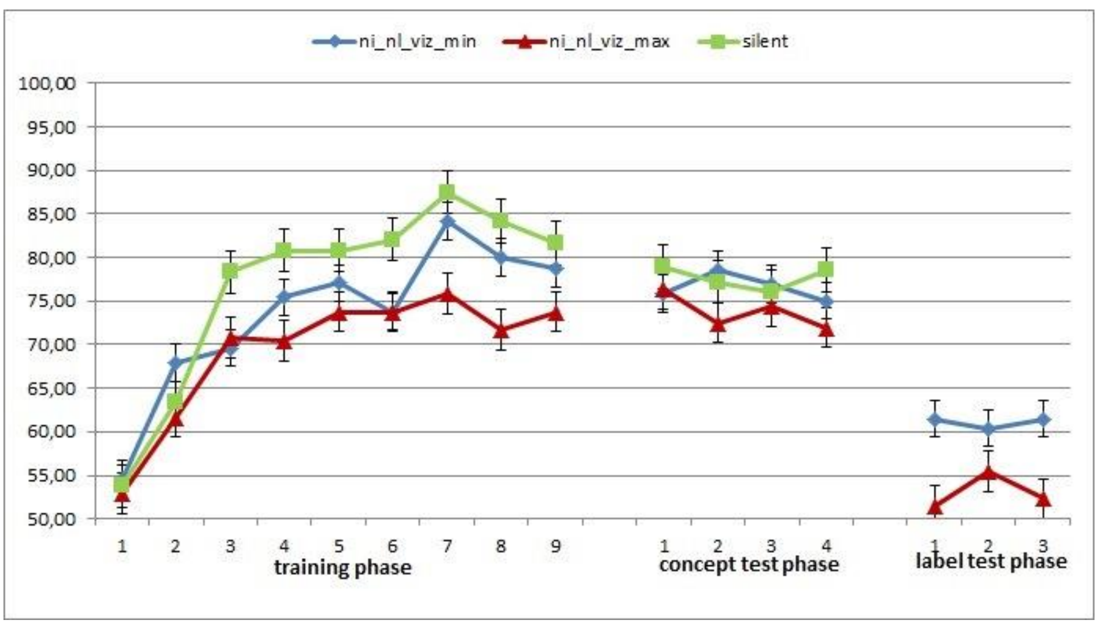

Figure IV-7: Percentage of correct responses over phases and blocks in Experiment 3

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor and Phase as a within-subjects factor with two levels (training and test).

There was no significant main effect of Group ($F(2,42) = 0.780$, $p = .465$, $\eta^2 = .036$). A main effect of Phase was marginally, but still not significant ($F(1,42) = 3.478$, $p = .069$, $\eta^2 = .076$). Finally, there was no significant Group x Phase interaction ($F(2,42) = 0.520$, $p = .599$, $\eta^2 = .024$).

In the label test phase, there was no significant difference between groups ($F(1,28) = 1.520$, $p = .228$, $\eta^2 = .051$).

These results follow a similar pattern as the results we found in chapter two: there are no effects of non-verbal labels on category learning. For further conclusions, a comparison between two conditions (with and without instruction) is needed.

Comparison with results from the experiment with instruction

A comparison between obtained results from this experiment and results from Experiment 1 from the first chapter will be conducted. These two experiments are completely the same, except the difference in the instructions given.

For *reaction times analysis*, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with four levels (minimum and maximum difference in both experiments) and Phase as a within-subjects factor with two levels (training and test).

Results showed that there was no significant main effect of Group ($F(3,56) = 0.344$, $p = .794$, $\eta^2 = .018$). However, there was no significant main effect of Phase ($F(1,56) = 0.583$, $p = .048$, $\eta^2 = .010$). Additionally, there was no significant Phase x Group interaction ($F(3,56) = 0.918$, $p = .438$, $\eta^2 = .047$).

Finally, there were no significant differences in reaction times in the label test phase ($F(3,56) = 0.406$, $p = .749$, $\eta^2 = .021$).

Results showed that participants responded with similar speed across conditions. This means that there was no speed-accuracy trade off. Concerning the difference in phases, it can be assigned to learning.

Accuracy. Descriptive results for all five experimental conditions are presented together in the following graph (Figure IV-8):

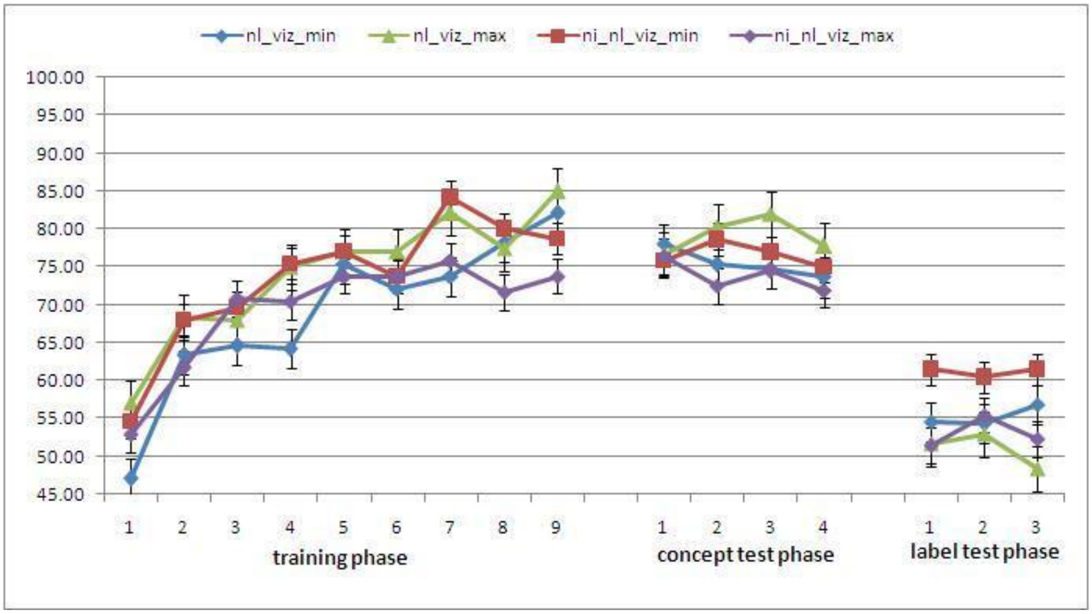

Figure IV-8: Comparison of percentage of correct responses over phases and blocks

As in the reaction times analysis, a two-way mixed-design ANOVA was conducted with Group as a between-subjects factor with four levels (minimum and maximum difference in both experiments) and Phase as a within-subjects factor with two levels (training and test).

There was no significant main effect of Group ($F(3,56) = 0.586$, $p = .627$, $\eta^2 = .030$). However, there was a significant main effect of Phase ($F(1,56) = 14.814$, $p < .001$, $\eta^2 = .209$). Finally, there was no significant Group x Phase interaction ($F(3,56) = 0.310$, $p = .818$, $\eta^2 = .016$).

In the label test phase, there were no differences between groups ($F(3,56) = 0.797$, $p = .501$, $\eta^2 = .041$).

Again, results showed that there were no effects of all types of visual non-verbal labels on category learning, no matter if there is experimental instruction or not. Participants learned labels similarly; however, this did not influence category learning.

5. DISCUSSION

In this chapter, the effects of instruction on labelled category learning were tested. Results showed that instruction can affect attention and also the learning of labels, which consequently can lead to results which depend on label type and modality.

Auditory verbal labels do not lead to better category learning when there is no instruction in the experiment. Additionally, the pattern of learning and the generalization of maximally different label condition and silent condition are similar, while minimally different condition is learned slower, but generalized similarly as in the other two conditions. This effect could be assigned to learning inhibition of minimally different label conditions, rather than the facilitation of maximally different condition.

Compared with those results obtained in the experiments with instruction (Chapter I), results showed that there is no difference between conditions, but there is a significant effect of interaction. This interaction is induced by a maximally different condition without instruction which is learned better than minimally different conditions, but generalized the same as those two. Furthermore, there is no difference between learning and generalization in this condition, while it is present for other three.

Written verbal labels also do not bring better learning or generalization. Unlike auditory verbal labels, there was no interaction between the factors. Finally, non-verbal labels do not have effects on category learning, no matter if there is instruction or not.

For all three types of labels, expectations from the introduction section were confirmed: labels do not lead to the effects in learning and generalizing when instruction that they need to be learned is omitted. But before we proceed with further interpretation, it is necessary to check the assumption whether we managed to manipulate attention on labels during the experiment. We were not sure whether participants would pay less attention to labels when there is no instruction. After obtaining the results, which showed that there are effects on learning while other experimental conditions were the same, we can conclude that the attention on labels was lower compared to the experiments where instruction was present. Further, those effects will be specified.

The most interesting effects were obtained with *auditory verbal labels*. Concerning the *learning*, we identified that maximally different labels without instruction did not have effects on category learning, unlike maximally different labels with instruction. We can assign these effects to lower attention on verbal labels which was induced by instruction. This result is in line with James' hypothesis.

Concerning a minimally different condition without instruction, learning was slower than in silent and with a maximally different condition. This means that participants paid some attention to labels and that these labels brought to an inhibition to category learning. Additionally, a minimally different condition without instruction was not different from a minimally different condition with instruction This signifies that minimally different labels could lead to a learning inhibition, no matter if there is instruction or not.

This is in line with James' hypothesis: once there are more similar elements, categories are learned slower. Minimally different labels represent another highly similar element. The problem is why participants still paid attention on minimally different labels, when they did not on maximally different labels? There is a possibility that these labels are less distinctive, which brings attention of the participants to identify differences. On the other hand, maximally different labels are, at the first glance, different, and are not further inspected, since they were task redundant.

Concerning the *generalization*, maximally different labels without instruction did not lead to better generalization. For the similar reasons like in learning, we presume that these effects are due to lower level of label learning, which decreases generalization once they are learned. This is in line with Lupyan's view, but also in line with James' hypothesis.

As far as minimally different labels without instruction are concerned, they also do not lead to better generalization, but this does not differ from maximally different or the silent condition. We can presume that labels do not allow for better generalization, since they are not learned well and hence do not differ from the no-label condition. This is in line with Lupyan's view, and also with James' hypothesis.

The obtained results fit into the previously proposed C*ategory learning based on the difference level model*. According to the model, the probability of labelled category learning is calculated with the following equation:

$$\mathrm{P(I,J)} = \mathrm{diff(I,J)}_T = \prod_{i=1}^{m} S_i * D_i * \prod_{i=1}^{3} S_i * D_i$$

Meaning, the total difference between categories relies on the level of visual features difference and label difference, but also on their attentional weights. Since in this experiment only attentional weights of labels were manipulated, we will interpret this section.

If label attention weight is higher, this leads to higher discriminability and a higher probability that categories will be learned. This is the case when labels are maximally different and the instruction is present. Once there is no instruction, categories will be less discriminable and consequently learned slower and generalized poorer.

In the case of minimally different labels, due to the nature of the labels, the absence of instruction does not lead to lower attention compared to maximally different labels without instruction. This makes learning and generalization similar to the minimally different condition with instruction. However, since labels are minimally different and adhered to visual features, this discriminability is lower, which leads to slower learning in minimally different condition.

The only remaining question is why generalization for minimally different condition does not differ from the silent one when labels are attended? The possible answer is that a learned category relies only on visual features, but not on labels. Even attended, these labels are not learned better and that is why their effects are diminished. Further possibility is that, as maximally different labels in Experiment 1 reached ceiling effects, we have here a kind of a bottom effect, where generalization cannot be lower.

As previously stated, the effects of written verbal and non-verbal labels were not identified, no matter if there was instruction or not. As far as *written verbal labels* are concerned, apart from the reasons stated in the previous chapter, we can presume that there is higher attention across modalities (for example, Allport, Antonis, and Reynolds, 1972). Once labels are written (meaning visual) they are in the same modality as visual features. The problem with this explanation is that labels and objects are not presented at the same time, which means that there is a lower possibility that they will interfere. However, it partially remains unclear why written verbal labels did not produce effects which were obtained with auditory verbal labels.

As far as *non-verbal labels* are concerned, their effect on category learning is the same with or without instruction. In the computational model in Chapter II it was specified that these labels represent only one dimension, unlike verbal labels which represent at least three. When attention is higher (not extremely high!), the value of this one dimension will not significantly contribute to the category difference as a multi-dimensional verbal label would. Hence, categories will not be learned faster and generalized better. This effect is even more significant once non-verbal labels are less attended.

C H A P T E R V

EFFECTS OF LABEL DIFFERENCES ON CATEGORY LEARNING – CONNECTIONIST MODELLING APPROACH

1. INTRODUCTION

In Chapter I, effects of verbal label differences on category learning were tested. In the conclusions of that chapter, it was stated that there are indications that the relationship between the level of label differences and their effects on category learning might be linear. Support to this conclusion was the effect of middle different labels, which lay between the effects of maximally and minimally different labels. It was also stated that testing this relationship would be very difficult with behavioural experiments, since it would be necessary to create several (at least ten) difference levels and to conduct experiments for each of them. Additionally, samples in those experiments should be significantly higher than in the experiments we conducted, for it would be very difficult to reach any significant level for small difference effects of labels.

However, there is a method to solve this problem and to get an understanding of the relationship between the level of label difference and their effects on category learning. That method is connectionist cognitive modelling. If we create a neural network which can demonstrate our findings, we can manipulate the level of difference between labels and test how many epochs are necessary to train the network to discriminate two categories. In this way, we can explore the functional relationship between label difference and the effects it has on category learning, which could be linear, exponential or logarithmic (chapter I). Additionally, a neural network that can demonstrate results obtained in the previous experiments that could lead to the conclusion what are the internal (in the brain) relationships between words and concepts.

It was stated that it “could lead to the conclusion” for a reason, since one can never be sure whether things really work that way. In connectionist modelling, there is always the possibility of ‘creating the network that works the way we want’. The aim of this chapter is not to create the model that will be the same as the one in the brain, but which could present “one of the possibilities”. Additionally, there is a problem of reductionism. Once we explain behavioural data with a connectionist network, we are moving from behavioural to a deeper level. Even though this approach is thought provoking and can be often revealing, it would be wrong to consider it as the only possible explanation of the behavioural data.

Some of the first connectionist models of semantic memory (in psychology) were the ones developed by Hinton (Anderson & Hinton, 1981) and by Rumelhart (Rumelhart, McClelland & the PDP Research Group, 1986; Rumelhart, 1993; McClelland & Rogers, 2003). The influence of the latter author was far more widely accepted.

In order to simulate Quillian’s hierarchical model, Rumelhart developed a model of association between concepts (plants and animals) and their attributes. The network consisted of three layers: representation, hidden and attributes. An additional layer (on the representation level) was the relation (isa, is, can, has).

After the training, a network successfully demonstrated relations between attributes and concepts. Additionally, the model demonstrated similar effects after it was damaged (as in the process of dementia), when essential differences between concepts were much longer preserved compared to specific attributes of the concepts.

Further development of this model was proposed by McClelland as the *complementary learning system theory* (McClelland, McNaughton & O'Reilly, 1995; McClelland & Rogers, 2003). This model predicts the existence of two learning systems; fast-learning and slow-learning. The former is hippocampal and occurs sometime after only one trial, while the slow-learning system is cortical and occurs at a slow pace.

There were some other connectionist models of semantic memory and learning, such as the ALCOVE model, developed by Kruschke and Nosofsky (Kruschke, 1992; Nosofsky, Kruschke, McKinley, 1992). This model demonstrated exemplar based category representation and learning, the model which was described in the introduction of this dissertation.

However, for our topic, the most relevant are the models that include verbal labels and demonstrate their effects on category learning. In last couple of years, there is a growing interest in cognitive models describing the effects of verbal labels on categorisation in childhood (Capelier, Twomey &Westermann, 2016, Twomey & Westermann, 2016, Twomey & Westermann, 2018). One such model was mentioned in the introduction chapter: the dual-memory model developed by Westermann and Mareschal (Westermann & Mareschal, 2014). The model predicted two types of memory: the fast-learning hippocampal system and the slow-learning cortical system. The model consisted of a three layered neural network, with an input, and a representation layer, which consisted of two parts: cortical and hippocampal (mutually interconnected) and an output layer from both systems (Figure V-1). Finally, the model included the task/ label units, which were labels of the category.

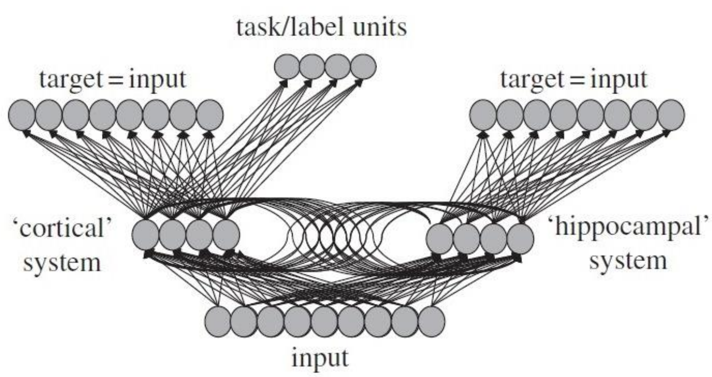

Figure V-1: Dual memory model (from Westermann & Mareschal, 2014)

Once categories are labelled with the global level labels, two principal components from the cortical representation units show much better differentiation between the categories, compared to non-labelled ones. Furthermore, once the verbal input is turned into an output, results show that labels are not just an additional feature, but rather object markers in a specific way: they do not make objects more similar, but representations of objects which are labelled are getting more prototypical, which is induced by label representation. In other words, labels are incorporated in object representation and are part of a so called **compound representation** (Twomey & Westermann, 2016; Capelier-Mourguy, Twomey & Westermann, 2016).

On the other hand, connectionist models that simulate effects of labels on categorisation in adults are significantly less frequent. One of the rare models is one developed by Gary Lupyan with which he demonstrated his Language augmented thought hypothesis Lupyan (Lupyan, 2012a).

The model consisted of three layers: perceptual, conceptual (representation) and a label level. All these levels were interconnected in two ways (Figure V-2). This interconnection could activate perceptual and label levels both internally and externally. The model learned two types of categories: goodies and baddies, which had some common features and some category specific features (which were produced by the function of probability).

In the learning phase, the model firstly learned *naming trials*, where the inputs were the visual features and output the name of the category. On the *comprehension trials*, the model needed to output visual features based on the name of the category and finally, on naming + comprehension trials, the model needed to produce both names and visual features, when provided with both.

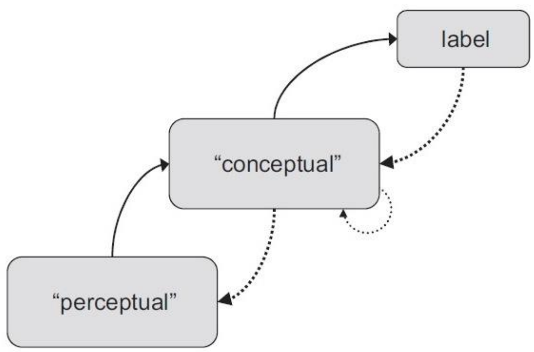

Figure V-2: Language augmented thought cognitive model (from Lupyan, 2012a)

In testing, firstly, the labels were disconnected from the representational level, so the model could reproduce a word, but it could not feed-back to the conceptual level. In the second type of testing, the feed-back from labels was included, but labels needed to be reproduced (they were not externally provided). In the final type of testing, labels were provided externally. In all testing trials, weights between nodes in the network were frozen.

Once labels are prevented from feeding back, the model significantly less effectively classified novel stimuli. This could be seen from the conceptual layer activation after principal component analysis. Based on this, it could be concluded that there are strong effects of labels on a representational level, which strongly goes in favour of the language augmented thought hypothesis.

The main aim of this chapter is to develop the model which will demonstrate label differences on category learning. Meaning, the aim is to demonstrate effects obtained in the experiments from the first chapter of this dissertation, where the effects of the phonological difference of verbal labels are recorded. Furthermore, this model should be able to demonstrate functional dependence between label difference and category learning.

For that purpose, a feed-forward neural network was created. This network has three layers: input, hidden (representation) and an output layer. There are two types of inputs: one for visual features of the stimuli and the other for the labels. The back propagation learning algorithm was used, the most frequently used learning algorithm for psychological studies.

2. METHOD

In this model, the connectionist network attempts to demonstrate the effects of label difference on category learning. Nine different models of networks are designed and trained with label pairs of various levels of difference.

Network design

The basic network used for all models has three levels: input, representation and an output (Figure V-3).

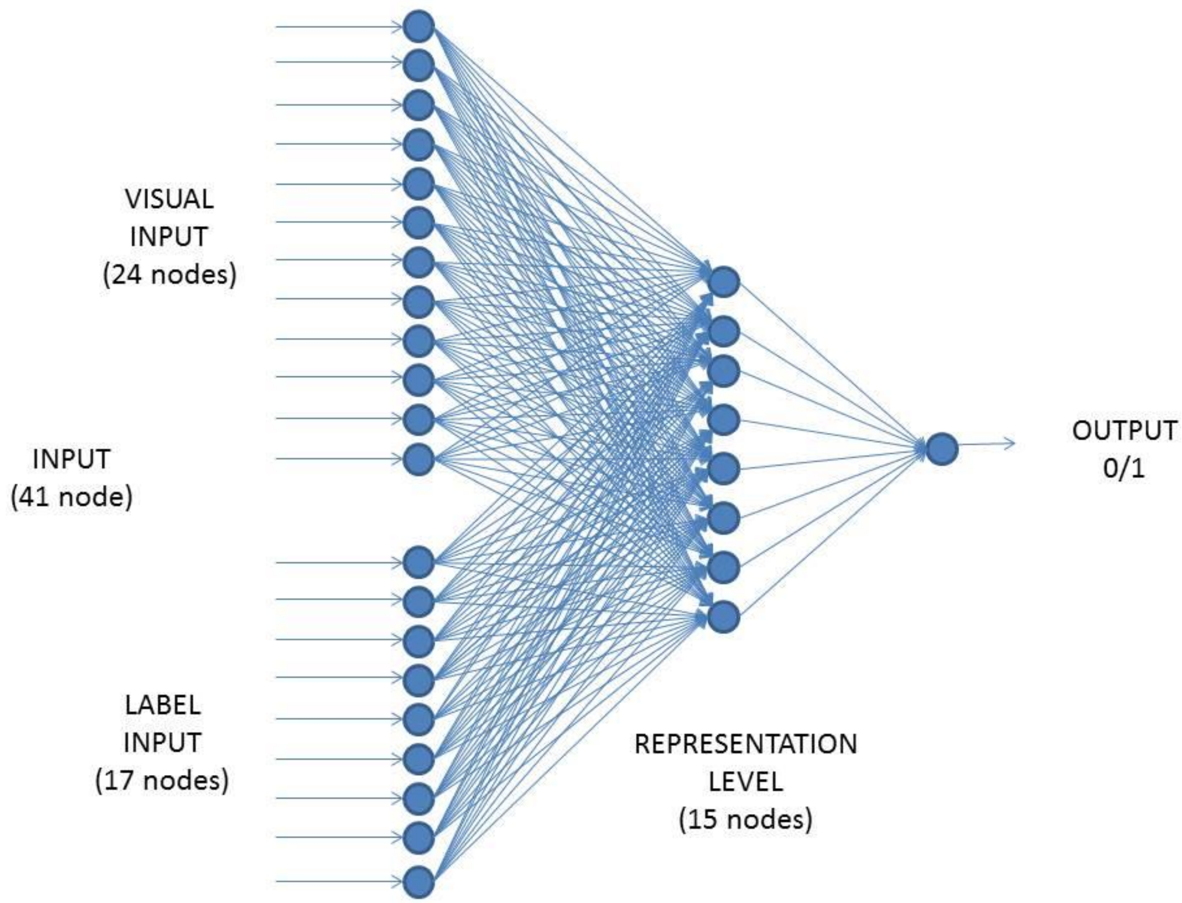

Figure V-3: Model of effects of label differences on category learning

Input level consists of two parts: visual stimuli and labels. Visual stimuli represents visual features of two categories properly coded (described below). It has 24 nodes. On the other hand, labels represent verbal input, which is also coded and has 17 nodes (also described below). There are 41 nodes in total in the input level.

Output level consists of a single node, which represents categorisation of the input (good or bad) with binary values (1 and 0 respectively).

Representation level consists of nodes which are in between input and output. This level is implemented as integrated representation (both for visual features and labels) and consists of 15 nodes. This makes the representation level compressed compared to visual or label input.

The number of nodes is determined by the number of inputs and outputs: primarily, it should be less than the number of inputs. Since there are 24 nodes of visual features, the number of representation level nodes should be lower than 24. Furthermore, the number of nodes should not be too low, since the total number of input nodes in most of the models is 41. Finally, since the output is a single node, too high a number of hidden nodes is not desirable. Taking all previous arguments into consideration, a number of 15 nodes is considered to be optimal.

Stimuli

Visual stimuli. Coding of visual stimuli contained the following elements: **common features** (8 nodes) and **category specific features** (16 nodes). Since YUFO stimuli used in these experiments differed on two features: the head and the base, each of these features are coded with 8 nodes.

Binary coding was used, where each node has a value 1 or 0. Category differences were coded as the following: for the first category (let us say "good"), the first four nodes of the feature are zeros (invariable), while the other four nodes are variable (for example – 0000.1101) in which within category specificities are coded. For the second category (let us say "bad"), the last four nodes describing visual features are invariable zeros, while the first four are variable (for example – 1011.0000). Likewise, within category differences there are 16 possible variations (2^4=16).

Common features are always coded as 10101010.

For example, let us take two members of the category of "good" and "bad" aliens:

Good: 10101010.0000.1101

Bad: 10101010.0111.0000

It is obvious that the values for the first eight nodes are always the same (10101010), while the following four of the "good" category are always zeros and the last four are within category specific values. It is the same for the "bad" category.

Complete values for 8 training tokens per each category and an additional 4 testing tokens (from Figure I-3), are represented in the following table:

Table V-1: Coding system for visual input

	Category of “good” aliens		Category of “bad” aliens	
Common features	Head	Base	Head	Base
10101010	0000.0001	0000.0001	0001.0000	0001.0000
10101010	0000.1010	0000.1010	1010.0000	1010.0000
10101010	0000.0011	0000.0011	0011.0000	0011.0000
10101010	0000.1100	0000.1100	1100.0000	1100.0000
10101010	0000.0101	0000.0101	0101.0000	0101.0000
10101010	0000.0110	0000.0110	0110.0000	0110.0000
10101010	0000.0111	0000.0111	0111.0000	0111.0000
10101010	0000.1000	0000.1000	1000.0000	1000.0000
10101010	0000.1001	0000.1001	1001.0000	1001.0000
10101010	0000.0010	0000.0010	0010.0000	0010.0000
10101010	0000.1011	0000.1011	1011.0000	1011.0000
10101010	0000.0100	0000.0100	0100.0000	0100.0000

Labels. One of the possibilities to code label differences is to use the three levels of differences used in Chapter I, where labels differed on three dimensions: the phonological structure of the label, sonority gradient and vowel position. Another possibility is to describe minimally and maximally different label pairs and to establish the scale between these two extremes.

The second approach will be implemented. In order to deeper describe the phonological difference between labels, based on linguistic measures, the following table of difference between maximally different labels (ketsi/ubom) and minimally different labels (dʒoset/ dʑoset) is constructed:

Table V-2: Quantified differences between maximally and minimally different labels

	Maximum Difference					Minimum Difference				
Label 1	K	E	ts	i		dʒ	o	S	e	t
Label 2		U	B	o	m	dʑ	o	S	e	t
Place	1		1		1	1				
Manner	1		1		1					
Voice	1		1		1					
Syllable-level	1				1					
Height		1		1						
Backness		1		1						
Rounding		1		1						
Total Diff					**17**					**1**

From the table, it is obvious that the maximal difference between labels is quantified as 17 and minimal as 1. Coding should satisfy this form, so the labels should differ maximally (at least) 17 points.

In order to construct such labels, 17 nodes are needed. If labels differ on each node difference it would be maximal (for example 00000000000000000 vs 11111111111111111). Minimal difference will be represented only on the difference on one node (for example 11111111111111111 vs 11111111111111110).

Since 9 difference levels are needed, coded labels with the following numbers are constructed: 1, 3, 5, 7, 9, 11, 13, 15 and 17. The values of the coded labels are given in the following table (with a highlighted difference of label 2 compared to label 1):

Table V-3: Coded labels and difference levels (the different bits in label 2 are highlighted)

Difference level	Coded label 1	Coded label 2
1	1.0101.0101.0101.0101	1.0101.0001.0101.0101
3	1.0101.0101.0101.0101	1.0100.0100.0101.0001
5	1.0101.0101.0101.0101	1.0110.0110.0101.0001
7	1.0101.0101.0101.0101	1.0110.0110.0011.0001
9	1.0101.0101.0101.0101	1.0110.0110.1011.0011
11	1.0101.0101.0101.0101	1.0110.1110.1011.0010
13	1.0101.0101.0101.0101	0.0010.1110.1011.0010
15	1.0101.0101.0101.0101	0.1010.1110.1010.0010
17	1.0101.0101.0101.0101	0.1010.1010.1010.1010

Procedure

Training/Learning. In order to simulate conditions of the experiments conducted in this dissertation, each of the trials within the epoch in network training will consist of two sub-trials: visual and visual + label.

A visual sub-trial will consist of the input from the visual nodes (24 nodes) plus invariable label input for all groups (0.5 for all label nodes). The latter is for the reason that participants do not hear the verbal label before the feedback, so the constant value of the label (0.5 for each node) simulates absence of verbal input. Since the participants in the experiment receive the feedback before they hear the label, after the visual sub-trial, the response will be checked and error correction propagated.

The second sub-trial consists of both verbal and label input. This sub-trial simulates exposition of the label to the participants in the experiment. An adult person can relate the label to the visual features, even though it was seen only after the feedback. After hearing the feedback and seeing/ hearing the label, adults can relate these two and this relation can be considered as a feedback. For this reason, error correction is propagated also after the second sub-trial.

The training lasts until the network successfully classifies all members within the epoch in the appropriate group (good or bad). Within each epoch, all stimuli are presented in random order. After all stimuli within a block are presented, the next block starts with the same stimuli.

In network training, each of 9 models was trained with the same (randomly selected) starting weights.

The function used in hidden layer nodes was tan-sig, while in the output layer nodes the function used was log-sig. The learning rate was adaptive and the scaled conjugate gradient algorithm was used. The network was trained and tested using the Matlab 2019a programme.

Testing. In testing, all weights within the network are frozen. Each model is tested for four additional stimuli per category (similarly as in behavioural experiment) and no label or feedback is presented. Learning is assessed by percentage of correct classifications.

Each of the models (for each level of difference) was trained and tested for the number of sets (N sets) and the average number of epochs were taken as a measure of network learning speed (for training) and the average number of successfully classified novels within category stimuli as a measure of classification accuracy (in testing). This method was used since stimuli are deterministic and easy to learn. Consequently, within model variability it may be present and possibly distort effects of label difference on network training in one set. A higher number of trainings will demonstrate relatively stable properties of the models.

The number of training sets (N) that were selected was 20, 40 and a 100. It is considered that N = 100 different training sets will be a sufficient number of sets for confident conclusions regarding the relation between two variables. On the other hand, N = 20 was selected, since there were 20 participants in the experiments in Chapter I. Finally, N = 40 was selected since it is a measure between the two extremes, which could indicate a possible trend on the scale of number of training sets.

3. RESULTS

Training

Training was conducted on different numbers of training sets (N=20, 40 and 100). Obtained results are shown in the following table (Table V-4):

Table V-4: Average training epochs for each model related to the difference level

Level of difference	Average of 20 sets	Average of 40 sets	Average of 100 sets
1	65.05	64.3	64.39
3	65.25	65.9	64.99
5	65.05	64.82	64.35
7	63.25	64.15	63.51
9	62.2	62.77	63.16
11	61.45	62	61.4
13	61.15	60.87	60.84
15	59.1	59.75	59.49
17	58.55	58.7	58.29

These results can be illustrated by the following plot (Figure V-4):

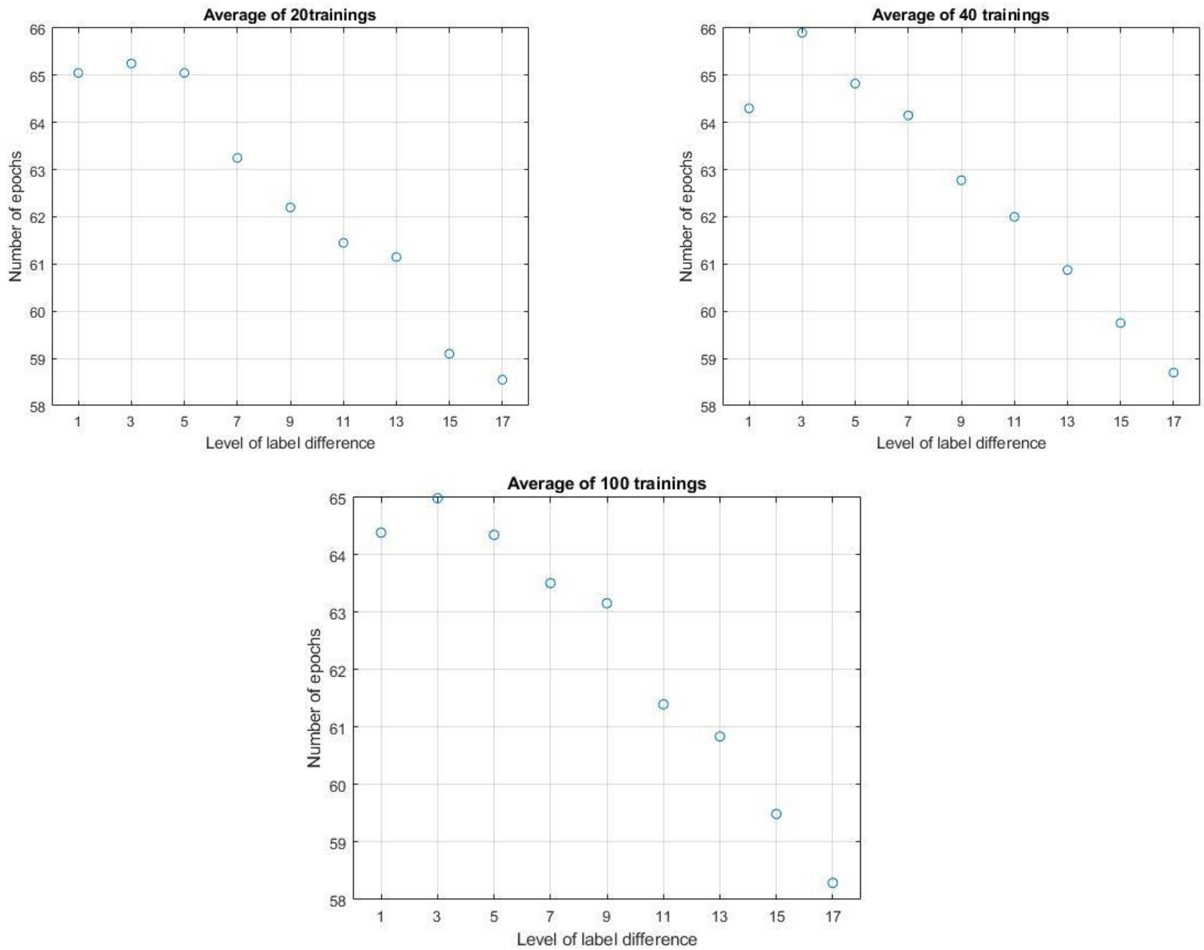

Figure V-4: Number of learning epochs for levels of label differences for 20, 40 and 100 training sets

As we can see from the obtained results, even if we change the number of training sets, the trend of results remains the same: small differences of labels lead to almost the same rate of learning categories. However, after a certain point (in this case difference of 7) a higher difference of labels, leads to easier learning of categories and this trend becomes almost linear.

Differences between the number of epochs is not high (maximum 8 epochs), but also, the number of learning epochs is small (as it was stated previously, for the reason that input properties are deterministic). Since the network was trained in a number of different sets (up to one hundred) and the trend was always the same, we can consider this trend valid.

The obtained learning rate induced by label difference can be used for approximation of the function of dependence between two variables. Using visual inspection, there are two probabilities of an underlying function: linear or polynomial (quadratic). There is a possibility to use the even polynomial cubic function, but there is a probability of over-fitting the data.

Approximation is conducted on the results from 100 training sets. Approximation of linear and quadratic functions of these data is presented in the following plot (Figure V-5):

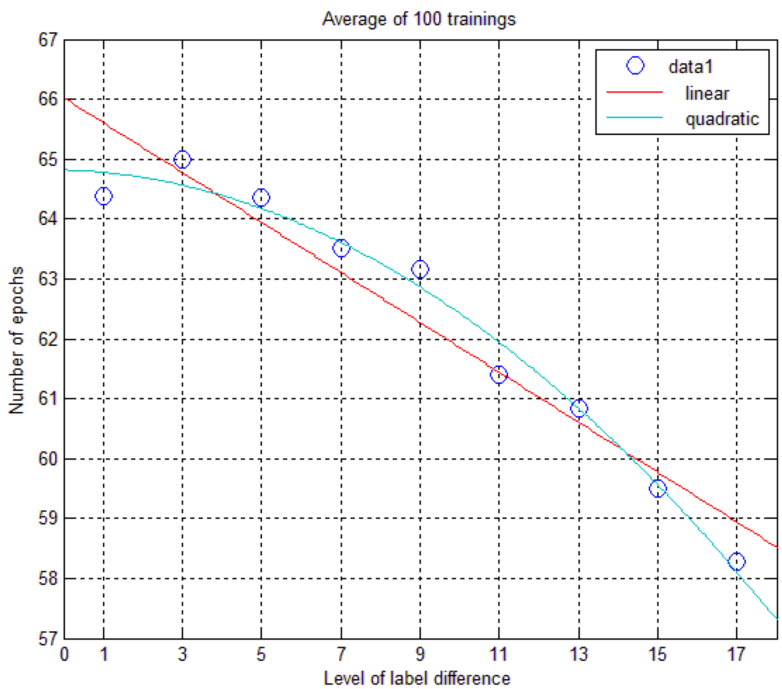

Figure V-5: Approximation function (linear & quadratic) of training epochs vs. label difference

Obtained functions from these two approximations are the following:

$$y = -0.42*x + 66 \text{ – linear function}$$
$$y = -0.02*x^2 – 0.019*x + 65 \text{ – quadratic function}$$

where *y* is the number of epochs and *x* is the level of label difference.

The norm of residuals from the linear function is 1.79, while for the quadratic function it is 0.89. This measure shows that the quadratic function approximates data much better compared to the linear function.

Residuals of these two functions are presented in the following chart (Figure V-6):

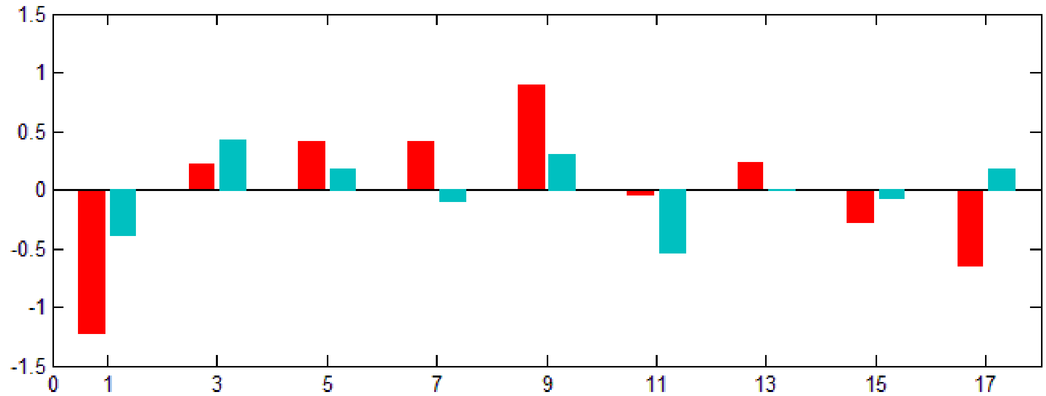

Figure V-6: Residuals of linear (red) and quadratic (blue) function

The obtained results, perfectly fit the data obtained in the experiments from Chapter I of this dissertation: as labels differ more, the effects on category learning are higher, meaning, categories are learned easier. However, this effect is not radical in the sense that small differences lead to huge consequences, but rather to the smaller but stable consequences. Since the sample in Chapter I was 20 students per group, it explains why there could not be an identified statistically significant difference between the middle difference and the maximal difference group, while there was a difference between minimal and maximal. It is highly probable that this significance would have been reached if the sample had been higher.

Testing

In network testing, each of 9 models were tested with new stimuli a number of times (N = 20, 40 and 100). For each of the models, the success rate was 100%, which is somehow expected, since stimuli were deterministic and easy to learn and generalize.

4. DISCUSSION

In this chapter, nine different models of neural networks were trained, each with the same visual input, but different label pairs with different levels of "phonological" difference. The network was trained in the manner to simulate input from the experiments conducted in Chapter I of this dissertation. The output was a single node representing classification of presented stimuli to one of the two groups (good and bad).

Results showed that the network learns faster once there is higher phonological difference of the labels, except for smaller differences between them. These learning rate differences are not enormous, but are steady and constant with an evident trend. These results perfectly fit the data obtained in Chapter I of this dissertation.

The interpretation of these results can be taken from different points of views: James' hypothesis, Lupyan's language augmented thought hypothesis and Westermann's compound representation approach. Additionally, we can interpret these results from the hypothesis developed in this dissertation: category learning based on the difference level.

From the point of view of *James' hypothesis* (James, [1931 (1890)]), which was the leading hypothesis of this dissertation, these results are completely expected: once labels are adhered to the concepts, and these labels are more different, the effects of learning will be evident. However, the problem with this interpretation is related to representation: implicitly it was considered that James' hypothesis considers the existence of two different types of representation: visual and verbal. Both of these representations are independent, which is not the case with the model we trained.

As far as Lupyan's *Language augmented thought hypothesis* is concerned (Lupyan et al., 2007; Lupyan, 2012a), a faster learning rate of concepts labelled with phonologically more different labels is not expected. The reason why labelled concepts are learned faster is because the labels are also learned in the process and they feedback activation of typical features of the concepts. No phonological differences between labels participate in such a process.

Additionally, Lupyan proposes (similarly like James) the independent existence of the representation of labels and concepts, which feed forward each other's activation. This is not the case in our model, since representations of objects (hidden layers in the network) are not separated.

Finally, there is a possible interpretation from Westermann's *compound representation approach* (Westermann & Mareschal, 2014; Twomey & Westermann, 2016; Capelier-Mourguy, Twomey & Westermann, 2016). Concepts are represented with labels together, not independently, which is what was obtained in this model. Even though Westermann never explicitly mentioned the possibility of effects of label differences, this is very probable: as much as categories (including labels) are different, hidden levels will represent those differences.

The final interpretation can be given from the position of *category learning based on the difference level model*. As labels are more different, the probability that labelled categories will be learned is higher. From Chapter I, this relation was expressed with following equation:

$$P\ (I,J) = diff(I,J)_T = diff(I,J)*diff(I_L,J_L)$$

meaning, the probability of learning two categories depends on the level of difference of not only their visual features, but also the difference between their labels.

This model does not presume whether representations of the labels are separated from representation of visual features or they are represented together. The model from this chapter amends this portion, demonstrating that these two representations are joint, or in Westermann terms, compound.

There are several consequences of this kind of interpretation:

1. All kinds of differences between inputs, whether it is visual or auditory, are relevant. The difference in label features is added up to the differences of visual features. This level of compound differences between categories is then relevant for category learning pace.

2. Possible differences between modalities are related to the modality or different coding induced by specific modality.

3. Representation level is not modality specific. No matter if visual features are visual and labels phonological, on a conceptual level, they are similarly coded.

Further interpretation of these results, will be presented in the conclusions of this dissertation.

Something that the model did not demonstrate, but which was demonstrated in the experiments in Chapter I, were the effects of label difference on *generalization*. All nine models successfully generalized novel stimuli with a 100% success, which was not the case in the behavioural experiments.

One of the reasons why this was the case could be the fact that the model learned up to the point when no mistakes were made in the entire epoch, which was not the case in the behavioural experiments. The training in the behavioural experiments was limited to nine blocks and the average success of the participants never reached 100%. Since training was not 100% successful, it is expected that test will not be 100% successful as well.

Finally, in this chapter, one major theoretical problem was not mentioned: the effects of non-verbal labels on category learning, even though this problem was analysed in this dissertation and it was predicted to be modelled. It appeared that it is a more difficult task than predicted, since it was not possible to successfully implement any of the previously mentioned theoretical backgrounds in the model and to demonstrate the effects obtained in the experiments with non-verbal labels. The closest to that was the model in which non-verbal labels were considered as moderately complex (presented with 9 input nodes) in which differences between labels did not show any significant difference in effects on category learning (which were obtained in the experiments in this dissertation). However, in this model, network learned categories faster compared to the ones modelled with “verbal labels”, which was not the case in the experiments. One of the reasons for these results might be that network learns faster once the number of inputs is lower.

It seems that this area will need more additional research and modelling to obtain satisfactory results.

GENERAL DISCUSSION

The aim of this chapter is to summarise results obtained in this dissertation. Furthermore, the aim, tasks, general and specific hypotheses stated in the introductory chapter will be recapitulated and individually elaborated. Additionally, general interpretation of the obtained results with existing and newly developed theoretical views will be integrated. Finally, in the last section, general conclusions of the dissertation will be stated.

1. SUMMARY OF THE OBTAINED RESULTS AND HYPOTHESES TESTING

Chapter I – Effects of verbal label differences on category learning

The ***primary problem*** of the first chapter of this dissertation was the following:

- Are the categories labelled with phonologically different words more easily learned compared to the categories labelled with phonologically similar words? This problem includes analysis of different level of similarity between verbal labels and its influence on concept formation.

Based on this problem, the ***general task*** of the first chapter was to identify effects of the level of phonological differences of verbal labels on category learning, using behavioural and neuro-physiological measures.

Founded on previously elaborated theoretical perspectives, the ***general hypothesis*** of this chapter was the following: Learning of categories labelled with phonologically different verbal labels will be faster and more easily generalized compared to the categories labelled with phonologically similar labels.

The ***Specific hypotheses*** were defined in the first chapter, and included the following:

- Categories labelled with phonologically more different verbal labels are learned faster and generalized better.

This hypothesis was confirmed. Participants learned faster and generalized better categories that were labelled with phonologically maximally different labels, compared to those labelled with minimally different labels, as it was demonstrated in Experiment 1 of this chapter.

- There is a continuum of effects of phonological labels differences. Meaning, categories labelled with middle different labels will be learned faster and generalized better compared to those labelled with minimally different labels, but worse than those labelled with maximally different labels.

This hypothesis was not fully confirmed. However, there are some indices that this continuum exists: while there were significant differences between maximal and minimal condition, the middle condition was not significantly different from either of them. This could indicate that this difference lies somewhere in between the two conditions.

- Categories labelled with maximally phonologically different labels are learned faster and generalized better compared to the no label condition (silent). This effect is higher compared to the effects of minimally different labels.

This hypothesis was partially confirmed. Maximally different labels do have an effect on category learning and generalization compared to the no label (silent) condition. However, this was not the case for minimally and middle different labels. Additionally, it is an important finding that obtained interaction in the Experiment 1 indicates that labels for the last two conditions do not have effects on learning, but do have effects on generalization. Meaning, labelled categories with minimally and middle different labels were not learned better compared to the no labelled categories, but were generalized better compared to them.

- Effects of label differences can be assigned to phonological difference, rather than to sound symbolism, even though the effects of sound symbolism could also be identified.

This hypothesis was confirmed. In Experiment 2 of the first chapter it was demonstrated that there were no significant differences between maximally different labels conditions with or without sound symbolism. Even some previous research produced these differences, we could not identify them, probably due to the "ceiling effect". Furthermore, it was demonstrated that phonological difference has an independent effect on category learning and generalization compared to sound symbolism.

Concerning the ***general hypothesis***, we can conclude that the hypothesis was confirmed. Learning of categories labelled with phonologically different verbal labels was faster and generalization was better compared to the categories labelled with phonologically similar verbal labels.

Chapter II – Effects of non-verbal labels on category learning

The ***primary problem*** of the second chapter of this dissertation was the following:

- Do non-verbal labels (visual and auditory) with different levels of similarity have effects on category learning?

Based on this problem, the ***general task*** of the first chapter was to identify the effects of the level of differences of non-verbal labels on category learning, using behavioural and neuro-physiological measures.

Founded on previously elaborated theoretical perspectives, the ***general hypothesis*** of this chapter was the following: Learning of categories labelled with phonologically different non-verbal labels will be faster and easier generalized compared to the categories labelled with phonologically similar labels.

The ***specific hypotheses*** were defined in the second chapter, and included the following:

- Categories named with maximally different visual non-verbal labels will be learned faster and generalized better compared to those labelled with minimally different non-verbal labels.

This hypothesis was not confirmed. While there were the effects of label differences on category learning with verbal labels, these effects were not present with visual non-verbal labels.

- There is an overall effect of maximally different visual non-verbal labels, compared to the no label (silent) condition.

This hypothesis was also not confirmed. Maximally different visual non-verbal labels do not lead to better learning or better generalization compared to the silent condition.

- Categories named with maximally different auditory non-verbal labels will be learned faster and generalized better compared to those labelled with minimally different non-verbal labels.

This hypothesis was not confirmed. While there were the effects of label differences on category learning with verbal labels, these effects were not present with auditory non-verbal labels.

- There is an overall effect of maximally different auditory non-verbal labels, compared to the no label (silent) condition.

This hypothesis was also not confirmed. Maximally different visual non-verbal labels do not lead to better learning or better generalization compared to the silent condition.

Concerning the ***general hypothesis*** of this chapter, we can conclude that this hypothesis was not confirmed. Non-verbal labels do not have any statistically significant effect on category learning and generalization.

Chapter III – Relation between effects of verbal and non-verbal labels on category learning

The ***primary problem*** of the third chapter of this dissertation was the following:

- What is the relationship between the effects of verbal and non-verbal labels on category learning?

Based on this problem, the ***general task*** of the third chapter was to identify relations between the effects of verbal and non-verbal labels on category learning, using behavioural and neuro-physiological measures.

Founded on previously elaborated theoretical perspectives, the ***general hypothesis*** of this chapter was the following: Linguistic and non-linguistic labels (visual or auditory) have the same effects on category learning.

In order to check this hypothesis, a comparison of the effects obtained in chapter I and chapter II needed to be tested. Prior to this comparison, the effects of the written verbal labels on category learning needed to be measured.

The ***specific hypotheses*** defined in the third chapter were the following:

- Categories labelled with phonologically more different written verbal labels are learned faster and generalized better.

This hypothesis was not confirmed. Unlike the effects obtained for maximally different auditory verbal labels compared to minimally different labels, written verbal labels do not differ regarding the effects on category learning.

- Categories labelled with maximally phonologically different written verbal labels are learned faster and generalized better compared to the no label condition (silent). This effect is higher compared to the effects of minimally different labels.

This hypothesis was also not confirmed. Unlike with the auditory verbal labels, learning categories with written verbal labels do not differ from the non-labelled ones.

Finally, related to the ***general hypothesis*** of the chapter, only auditory verbal labels have effects on category learning. In this sense, categories labelled with maximally different auditory verbal labels are learned faster and generalized better compared to other types of non-verbal labels.

Chapter IV – Effects of attention on labels induced by experimental instruction on category learning

The ***primary problem*** of the fourth chapter of this dissertation was the following:

- Whether the attention on labels induced by experimental instruction can produce effects on category learning and generalization? This problem also includes the question whether phonological difference of labels can have different influences on category learning?

Based on this problem, the ***general task*** of this chapter was to identify the effects of experimental instruction on category learning, using behavioural measures.

Founded on previously elaborated theoretical perspectives, the ***general hypothesis*** of this chapter was the following: Experimental instructions in which it is explicitly required from the participant to learn the label will have a higher effect on category learning, compared to the experiment in which there are no experimental instructions.

The ***specific hypotheses*** defined in this chapter were the following:

- More attended auditory verbal labels will have higher effects on category learning and generalization.

This hypothesis was confirmed for maximally different labels. Once there is no experimental instruction, the effects of maximally different verbal labels are not recorded. Effects for minimally different verbal labels are reflected in the fact that there is an inhibition in the learning, but not in the generalization process. Furthermore, effects of maximally different labels without instruction on learning are the same, both in learning and in generalization compared to a non-labelled condition.

- More attended visual verbal labels will have higher effects on category learning and generalization.

This hypothesis was not confirmed simply because more attended visual verbal labels did not produce any effects on category learning.

- More attended non-verbal labels will have higher effects on category learning and generalization.

Similarly as with the previous hypothesis, this hypothesis was not confirmed, because more attended non-verbal labels did not produce any effects on category learning. Likewise, similar effect was recorded with non-attended ones.

Concerning the ***general hypothesis***, we can conclude that the hypothesis was confirmed. Learning and generalization of categories with explicit instruction that labels need to be learned (more attended) is better compared to the case where this instruction is omitted (less attended labels).

Chapter V – Effects of label differences on category learning – connectionist modelling approach

The ***primary problem*** of the fifth chapter of this dissertation was the following:

Is it possible to create cognitive models which will be able to demonstrate obtained results from the previous experiments? Furthermore, is it possible to reveal functional dependencies between levels of label difference on category learning?

Based on this problem, the ***general task*** of the first chapter was to create physiologically plausible cognitive models which will demonstrate the effects of label differences on category learning.

Founded on previously elaborated theoretical perspectives, the ***general hypothesis*** of this chapter was the following: It is possible to create physiologically plausible cognitive models which will demonstrate effects of label difference on category learning.

The **specific hypotheses** defined in this chapter were the following:

- A developed connectionist model can demonstrate effects of label differences on category learning.

This hypothesis was confirmed. The model that was developed in Chapter V does demonstrate effects of label differences on category learning. Once labels are more different, categories are learned faster compared to categories labelled with less phonologically different labels.

- A developed connectionist mode can record functional dependence between label difference and category learning.

This hypothesis was also confirmed. In Chapter V the functional dependence between different levels of label difference and category learning was demonstrated. This interdependence is rather quadratic than linear, but also a linear function can explain much of the variance of the model.

- A developed connectionist model can demonstrate different effects of verbal and non-verbal labels on category learning.

This hypothesis was not confirmed. Since this problem was more complex than expected, work on this was left for further research.

Concerning the ***general hypothesis***, it can be considered to be confirmed, since it was a developed cognitive model which successfully demonstrated effects of verbal labels on category learning.

2. GENERAL INTERPRETATION OF THE OBTAINED RESULTS

In this section, the obtained results from previous chapters will be interpreted from the point of view of different theoretical perspectives. These interpretations will be separated for behavioural and connectionist modelling results. In the final section, these interpretations will be integrated.

Concerning the behavioural results, the following perspectives will be analysed: James' hypothesis, Lupyan's language augmented thought hypothesis, cognitive development approach (Waxman's view of labels as conceptual markers and Sloutsky's view of labels as object features). Finally, the alternative category learning based on the difference level approach will be presented.

Concerning the connectionist modelling results, the results will be interpreted from the perspectives of James' hypothesis, Lupyan's language augmented thought hypothesis and Westermann's compound representation approach. Again in the end, results will be interpreted from the point of view of Category learning based on the difference level approach.

INTERPRETATION OF BEHAVIOURAL RESULTS

James' hypothesis

According to James' hypothesis (James, [1931 (1890)]), once categories are adhered with more distinct elements, these elements drag those categories apart and make them more distinct. Thus, it is easier to learn them and consequently easier to generalize from them. As we saw from the Introduction and Chapter I, these elements could be many different things, including labels. Once categories are named, since names are more distinct, categories will be learned faster and generalized better.

This hypothesis perfectly explains results obtained with auditory verbal labels. If verbal labels are maximally different, they lead to faster learning and better generalization of categories. However, categories labelled with minimally different labels, are learned more difficult than those labelled with maximally different labels. It is expected that there is a continuum between phonological difference of labels and category learning and generalization.

The problem with James' hypothesis is related to the effects of non-verbal labels on category learning. This hypothesis was not capable to explain these results. According to James' hypothesis, there is nothing inherently different between verbal or non-verbal labels, while they are distinct enough to bring category differences to a higher level. However, results from these experiments proved differently: verbal labels (auditory) do have effects on category learning and generalization, which is not the case with non-verbal ones and more surprisingly, with written verbal labels. As we noted previously, some authors (Waxman & Gelman, 2009; Lupyan, 2012a) claimed that words have a special status.

Generally, James' hypothesis was the basic hypothesis of this dissertation. Since this hypothesis fails to give a full explanation of the phenomena described in this dissertation, we need to look for alternative explanations.

Lupyan's language augmented thought hypothesis

As it was noted in the introduction, Lupyan's language augmented thought hypothesis considers mutual interdependence between thought and language (Lupyan, Rakison & McClelland, 2007; Lupyan, 2012a; Lupyan, 2012b; Lupyan, 2015; Edmiston & Lupyan, 2015). Concepts are connected to labels in the long term memory. Once a concept or a specific exemplar is activated, it activates the labels, which in return loops back and activates typical features of the category. In that way, the most typical elements of the category are activated, which makes category a more "categorical". Additionally, concerning the effects of verbal labels compared to non-verbal labels, Lupyan considers verbal labels as unmotivated cues which are related to all category members, unlike some non-verbal motivated cues which are related to some specific subcategories.

Lupyan's hypothesis explains well the way in which labelled categories are learned, which was his main hypothesis in his seminal work (Lupyan et al., 2007). Simply, labelled categories are learned faster, since in the process of learning labels, they increase typicality of the category and makes identification easier of typical features of the category.

Furthermore, Lupyan explains well the process of generalization. As it was noted, labels activate the most typical features of the categories. In that way, labelled categories will be generalized better, compared to the ones which are not labelled.

There are however, some findings in this dissertation which cannot be explained by Lupyan's hypothesis. In the first place, how maximally different labels bring to faster learning and generalization compared to minimally different labels? Lupyan's hypothesis does not recognize the possibility that the level of label difference can influence learning and generalization of the categories.

According to Lupyan's hypothesis, every label, no matter if it is different to the other label or not is equally effective in boosting categorical salience. Two categories which have either highly different or highly similar labels will be equally learned and generalized. However, findings in this dissertation, especially those in the first chapter, proved the opposite.

Additionally, one notable problem is the difference between effects of verbal and non-verbal labels. According to Lupyan, once these labels are unmotivated cues, there should be no difference in their effects on category learning and generalization (Edmiston & Lupyan, 2015).

However, in this dissertation, we proved differently. In Chapter II, we demonstrated that non-verbal labels, even if they were unmotivated cues (labels did not vary with some elements of the categories), effects on category learning and generalization were not identified. It seems that there is a special status of words in human memory, but this status should be assigned to some other property, rather than its unmotivated nature.

In the further text, we will try to additionally examine the nature of verbal labels compared to the non-verbal ones. Unfortunately, there are not many theories which were developed for adults in this area. Lupyan's language augmented thought is one of the rare. For that reason, we need to turn to theories created in developmental cognitive psychology, hoping that effects they describe are universal and can be applied for adults too.

Cognitive development approach

As it was stated previously, this approach will be split in two basic views: Waxman's view of labels as conceptual markers and Sloutsky's view of labels as object features.

- ***Labels as conceptual markers***. As stated in the introductory chapter, this theory was developed by **Waxman** and her associates (Waxman, 1991; Waxman & Markov, 1995; Fulkerson & Waxman, 2007; Waxman & Gelman, 2009). This theory views labels as conceptual markers which refer to categories. In this sense, referring means that labels are related to more abstract category members, compared to any individual category member ever seen or which will ever be seen (Waxman & Gelman, 2009). This means that labels serve as an invitation to form categories.

Non-verbal labels do not have this kind of property. They do not serve as an invitation to form category. For that reason, once the categories are labelled with verbal labels, they are learned faster and generalized better.

We can only presume where this property of labels comes from, since Waxman does not specify the nature of this property. One of the possibilities is that we have inner instinct (or nature) to interpret verbal labels as something that marks categories. Implications of this view would be that there is a kind of a meta-categorical disposition in human brains, which is innate and which content is filled with specific experience.

This view is in a sense close to rationalistic philosophy, very close to the views of Plato, Descartes, Leibniz or Kant. Even though this view is very interesting and thought provoking, it opens additional questions, on which this theory does not give explicit answers, such as: where this verbal labels invitation property comes from and what is its nature?

This explanation is in a sense still scientifically valid. For example, we still do not know what in essence electricity is, but science developed an enormous body of knowledge related to the implementation of the electricity. This kind of theory gives us a good inspiration for empirical research which reveals many effects of verbal labels on concepts development and learning.

However, a problem with this kind of theories is that often the same results one can interpret differently. For example, once you say that verbal labels represent invitation to form category, you can interpret all results where effects of verbal labels on category learning are recorded, not giving a precise explanation how these effects work. Similarly, results in this dissertation could be interpreted this way. However, unless you are capable to demonstrate how these mechanisms are conducted, the given explanation is only partially valid.

Finally, there are some results obtained in this dissertation which cannot be explained by this theory, such us: how to explain the effects of verbal label difference level? If labels represent invitation to form categories, they would do it equally, no matter if they are phonologically different or not.

- ***Labels as object features view***. This view was developed by **Vladimir Sloutsky** and his associates (Sloutsky & Lo, 1999; Sloutsky & Fisher, 2004; Sloutsky & Fisher, 2012; Sloutsky, 2010) and it represents a coherent view of child concept development. While Waxman's model is in a sense an extension of rationalistic philosophy, this model strongly relies on empiricist tradition.

According to this view, labels are not category markers, but rather object features, just like any other feature. Once categories share the same name, this name contributes to their overall similarity, which leads to the more probable conclusion that the two exemplars are members of the same category.

Sloutsky further specifies his claims in his **SINC** (Similarity Induction Naming Categorisation in young children) model (described in the introductory chapter), where he calculates the probability of categorising a certain exemplar in a given category. Probability will be higher, if the exemplar is more similar to the other members of the category. By using multiplicative rule, some very different elements could lead to very high differences between the given exemplar and the other category members.

Sloutsky's model explains well both learning and generalization in our experiment, as far as verbal labels are concerned. If the verbal label is another feature contributing to overall similarity, than it will be easier to learn labelled compared to the non-labelled categories. Exactly this was obtained in our experiments.

Furthermore, Sloutsky's model explains well the effects of label differences. In his model he predicted that similarity of the labels (meaning phonological similarity), lead to a higher similarity of the labelled categories and higher possibility that they will be estimated as members of the same category.

There are, however, some shortcomings of Sloutsky's view: if a verbal label is just another feature, like any other non-verbal feature, the question is why there are effects of verbal labels, but no effects of non-verbal ones, which is also obtained in our experiments? Furthermore, why there are effects of label difference on category learning labelled with verbal but not on non-verbal labels.

According to Sloutsky, the reasons for these effects are higher attentional weights of auditory labels. Auditory labels bring to the higher attention and easier learning of the categories. However, the problem with this proposition is that we did not obtain any effects of auditory non-verbal labels either, which means that there is something inherent in the non-verbal labels compared to verbal labels, but not modality in which they are presented.

The final problem with Sloutsky's view, which can be also applied to all theories of cognitive development, is that they are created for developmental psychology. We can never be sure whether these models are universal for adults. Furthermore, Sloutsky himself showed different results for adult participants in his experiments. Even if we use an argument that elements which we describe in the child development approach are the same and universal, we can never be sure about it.

For this reason and for the reason to explain other results obtained in this dissertation which could not be explained by one theory or hypothesis solely, the model of category learning based on the difference level was developed, which will be further specified.

Category learning based on the difference level

We will start the description of this model with individual dimensions. For example, if we have two categories which differ on three relevant dimensions (example: colour, size and shape), each of these dimensions can be quantified and represented as a difference between categories. For example, the colour spectrum can be divided from white to black, where each of the shades can be quantified. Those numeric values would represent the value of each dimension of the category.

But the same category members might be of the different colours. In that case, the colour dimension for the category will be represented with the mean value of each of the members. In that case, the difference between the two categories A and B would be the following:

$$d_{between} = |mean\ A - mean\ B|$$

This difference represents the absolute value of the difference between the two categories on a certain dimension (for example colour). For the reason which will be described later, values of this difference are going from 1 (minimal difference or no difference) to 100 (maximal possible difference). In terms of colours, if all category members for both categories are white, this difference will be 1. If members of one group are white, and members of the other group are black, this difference will be 100.

However, this difference does not represent the total difference between two categories on the certain dimension. Variability within categories significantly contributes to the overall difference. For example, it is not the same if all members of the category share the same colour or not. This type of difference, we can name as within category difference and it will represent variability within the group which can be calculated by the following equation:

$$d_{within} = \frac{\sum_{i=1}^{n} |x_i - mean|}{n}$$

Where x_i represents the value of *i*-th category member on the dimension and *mean* is the mean of the category on the dimension, while *n* is the number of category members. Similarly as with between group difference, this difference can have values from 1 (minimal within group difference, or no difference) to 10 (maximal possible within group difference).

There are, however, two types of this difference: one within the first category and the other within the second category. Both of these differences, along with between group difference contribute to the overall (or total) group difference, but in different directions: while higher between group difference leads to a higher total difference, higher within group difference leads to a lower total difference. For example, more colours are within each of the categories, the overall difference between categories will be lower, since some colours may be similar or even overlap.

Having on mind this property, we can calculate total difference (D_i) on measured category dimension with the following equation:

$$d_{tot} = \frac{d_{between}}{d_{within1} * d_{within2}}$$

As we can see from the equation, the total difference on category dimension is directly proportional to between group differences and indirectly proportional to within group differences.

An important property of these differences is pondering. Between group differences are pondered on a scale from 1 to 100, while within group differences are pondered on a scale from 1 to 10. Why? The reason for this pondering is to avoid values from 0 to 1 which would go in the fraction numerator, since those values would lead to very high overall value. For example, if the one within group difference was 0.1 and the other was 0.2, multiplied this would be 0.02. Once we divide an even small value with this number (example 10), we would get a score of 500, which is an enormously high difference value.

Furthermore, the between group difference value is pondered to 100 maximal value, so the division between two maximal values (10*10 = 100 in the numerator) would give 1. Thus, we get a possible total difference variability from minimally 100/1 = 100 to maximally 1/100 = 0.01.

The story of pondering does not finish here. These values are also pondered so it can take values from 0 (minimal difference or no difference – equal to 0.01 non-pondered value) to 1 (maximal possible difference – equal to 100 non-pondered value). This pondered value of the total difference is marked with a capital D.

For example, in terms of colours, two categories will be maximally different on colour dimension if their within group members are of the same colour (no within group differences, meaning values are 1) and the maximal between group differences (black and white, value is 100). The score for the total difference will be d_{tot} = 100/1 = 100, which will after pondering get a value of D = 1, meaning the maximal possible difference between categories on the dimension of colour.

Not all dimensions are equal. Some are more important than others. For example, for differing a category of pencils from a category of pens, the difference in colours will hardly contribute much to their difference. In this case, shape might be more crucial than colour.

These dimensional differences can be represented with *f* – feature weight. Values of this feature weight vary from 0 (no weight at all) to 1 (maximal feature weight, meaning the only element, crucial for category membership).

Apart from feature weight, an important element is also something that we can name as attentional salience - s. For example, a head of a figure is usually more salient than a tail. Values for this salience go from 0 (no salience) to 1 (maximal salience which everyone notice at first). However, these values usually do not vary that high. For example, the typical value for a head would be 0.60 and for a tail 0.40.

The total weight of these two values could be expressed as S, or attentional weight, which is calculated as follows:

$$S = f_i \, * \, s$$

meaning that the attentional weight is equal to the product of feature weight and attentional salience of the individual dimension.

Using these two equations, we can express the level on which the individual dimension contributes to the between category differences:

$$\text{D} * \text{S}$$

With the above presented equation we can calculate differences between categories on the individual dimension. The problem is how to calculate the total between category differences on all relevant dimensions. One of the options is to use multiplicative rule for all dimensions. The problem is that once this rule is used, the low difference on one dimension, can lead to very low total differences. For this reason, we can use the summation rule.

Using the summation rule, we have two options: to use Euclidean distance or average value. Even though Euclidean distance has more power (it can represent correlations between dimensions), for practical reasons, we will use average value, since average value will keep results values between 0 and 1, which is needed for probability calculation.

Based on the previous claims, we can calculate between categories difference level – diff of the two categories A and B by the following equation:

$$diff\ (A, B) = \frac{\sum_{i=1}^{n} D_i * S_i}{n}$$

Where n is the number of relevant dimension on which two categories differ. By using this equation, since values vary from 0 to 1, we can calculate probability that categories will be learned (differentiated), where higher probability means faster learning:

$$P(A, B) = \ diff\ (A, B) = \frac{\sum_{i=1}^{n} D_i * S_i}{n}$$

This is the equation of the category learning based on the difference level. However, in light of the findings in this dissertation, one important element is missing and that is the label.

Similarly to the category difference, **label** difference level can be calculated, except that there are no within category differences, since one label is used for the entire category.

$$d_{L-tot} = |mean\ A - mean\ B|$$

This difference represents total difference on selected dimension. In order to calculate overall between labels difference, we can use a similar equation as for the categories:

$$diff\ (L1, L2) = \frac{\sum_{i=1}^{m} D_i * S_i}{m}$$

This equation shows that difference between two labels is calculated as an average value of difference and attentional weights on *m* relevant dimensions. As it was noted in Chapters I and II, this number may vary, and we presume that for the verbal labels it is at least 3, but for the non-verbal labels it is 1. This gives the difference between verbal and non-verbal labels and gives the verbal labels special status.

If we merge two equations, we get the following final equation:

$$diff\ (A, B)tot\ = \frac{m \sum_{i=1}^{n} D_i * S_i +\ n \sum_{i=1}^{m} D_i * S_i}{n * m}$$

And the probability of learning labelled categories could be expressed with the following equation:

$$P(A, B)tot\ =\ diff(A, B)tot = \frac{m \sum_{i=1}^{n} D_i * S_i +\ n \sum_{i=1}^{m} D_i * S_i}{n * m}$$

Based on this equation, we can estimate probability in which labelled category can be learned in a given time. We can see from the equation that the labels of the categories can contribute to the labelled categories difference in two ways: if they are more different than non-labelled categories, they influence to a higher difference between labelled categories. If they are less different than non-labelled categories, then they contribute to higher similarity between labelled categories. This is in line with James' hypothesis, that labels which are more different than categories, once are adhered to them, stretch those two further apart.

There is however, a problem with this model. We can have two categories which are very similar in almost all but one dimension, which is very salient and very different (let us say S = 0.9 and D = 0.9, while other dimensions are S = 0.1 and D = 0.1). In this case if the number of dimensions is high, the average difference will be lower. For example, if we have two groups of aliens, which are almost the same on all dimensions but colours (example: one group is yellow and the other is blue). Calculating the overall difference based on the previous formula would look like the following (S and D values respectively):

Colour: 0.9*0.9 = 0.81;
Head: 0.2*0.1 = 0.02;
Tail: 0.1*0.1 = 0.01;
Horns: 0.1*0.1 = 0.01

The average value of these dimensions would be:

(0.81+0.02+0.01+0.01) / 4 = 0.21

This is a very low difference level, but categories would be easily differed: it is just necessary to notice their difference in colour.

In order to avoid this situation, the easiest way is to turn difference levels on dimensions to similarity and then use the multiplicative rule:

$$sim_i = 1 - diff_i$$

where sim_i represents similarity on the dimension i. Once we use this equation, we can calculate overall similarity using the multiplicative rule, suggested by Medin and Schaffer (Medin & Schaffer, 1978):

$$sim\ (A,B) = \prod_{i=1}^{n} sim_i$$

Once this between categories similarity is calculated, we can turn it back to difference by the inverse equation to the previous:

$$diff\ (A,B)\ =\ 1 - \ sim(A,B)$$

And now, this difference can be used for learning probability calculation.

With this equation, calculation from our example would look like following:

$$sim\ (A,B)\ =\ 0.19\ *\ 0.98\ *\ 0.99\ *\ 0.99 = 0.18$$

After similarity, we can calculate back the total category difference:

$$diff\ (A,B) = 1 - 0.18 = 0.82$$

This difference, unlike the one used by summation rule looks more plausible since one dimension which has high salience and high difference, no matter if other dimensions are almost the same, contributes highly to the overall difference between categories.

Using similar method, we can calculate label total difference level. In that case, we primarily calculate the total between categories similarity (after dimensional differences are turned into similarities) including features and labels:

$$sim_T = sim\,(A,B) * sim\,(L1, L2) = \prod_{i=1}^{n} sim_i * \prod_{j=1}^{m} sim_j$$

And finally, this similarity is turned to difference, which in return equals to probability of learning category in a given time.

$$P(A,B)_T = diff(A,B)_T = diff(A,B) * diff(L_1, L_2)$$

Where $P(A,B)_T$ represents the total probability that labelled categories will be learned. As we can see, label difference also contributes to category difference. Once these labels have several dimensions (like we presumed for verbal labels), their contribution to overall categories similarity will be higher. If labels represent only one dimension (like we presumed for non-verbal labels), there will be their contribution, but not as high as the one from the verbal labels.

This model is capable to explain most of the findings in this dissertation. Prior to further elaboration, it is important to notice that this is speculative quantifying. Things might work this way, but also might not. For further confirmation, additional empirical findings would be needed. However, defined like this, the model explains well many of the obtained data related to effects of labels on category learning.

We can now list findings in this dissertation which this model can explain.

Concerning the **learning**, two labelled categories can be learned faster if the adhered labels are more distinct. Difference of labels, leads to higher overall difference of the categories, which in return brings to a higher possibility that they will be learned (meaning easier and faster). This is exactly what was obtained with auditory verbal labels. Once labels are less distinctive, they lead to higher similarity between categories, which can lead to slower learning and even inhibition compared to the no labelled condition.

Obtained effects cannot be implemented for the non-verbal labels, for the property which will be later discussed.

Concerning **generalization**, we can record similar effects as in the learning, except that we now talk about mental representations. Once categories are learned (they are concepts now), and when they are associated to the labels representation, the difference between labels is also appropriately represented. More distinct labels lead to easier discriminability between concepts once they are activated.

This view is close to Lupyan's view of mutual on-line activation of labels and category features. However, unlike Lupyan, this view considers that even representation of the labels remain its difference, which can lead to better discriminability between concepts and likewise better generalization.

Finally, concerning the **difference between verbal and non-verbal labels**, as we previously noted, these differences can be explained by the difference in complexity of the verbal and non-verbal labels. While verbal labels are more complex and are composed of several dimensions, non-verbal labels are one-dimensional. This complexity leads to higher weight of verbal labels, compared to non-verbal labels, which is manifested in category learning and generalization.

The nature of this complexity remains unclear. We demonstrated that there could be at least three dimensions on which verbal labels can be differentiated – phonological structure of the label, sonority gradient for alveolar/postalveolar sounds and vowel position. A possible additional dimension is sound symbolism. Apart from sound symbolism, previous dimensions are linguistic properties, but we cannot be certain about their psychological reality. It seems that there will be further research required to specify these verbal label dimensions and their cognitive representation, which makes them more complex compared to other types of labels.

This view explains where the special status of words comes from. Verbal labels are not unmotivated cues, as Edmison and Lupyan claimed, or they are not just another feature which is not different from other features, like Sloutsky claimed. Verbal labels are also not referring to the categories and invite for their formation, but rather another feature which is more complex and which effects are more notable compared to other labels.

However, one can question this kind of explanation: what tells us that non-verbal labels are one-dimensional? In Chapter II, non-verbal written labels were made using several variables, which opposes claims they were one-dimensional cues.

In order to further explain effects of verbal and non-verbal differences on category learning, we can include a constant l^* in our equation (l stand for labels and the asterisk signifies that it is a constant). This constant is related to the nature of labels: whether they are verbal or non-verbal, auditory or visually presented. Additionally, it is related with relevant experience of the individual with a specific label type.

The value of this constant goes from 0 to 1: if it is 0 (better to say close to 0), the effect of a specific type of label on the total category difference is very low. If this constant is close to 1, then the effects of a specific type of label on the total category difference is very high (close to their difference level).

Once we use this constant, our focal equation of probability of category learning becomes:

$$P(A,B)_T = diff(A,B)_T = diff(A,B) * l^*[diff(L_1,L_2)]$$

From the experimental results in this dissertation, it can be concluded that verbal labels auditorily presented have the highest l^*. Let us say that this value is 0.9 or even higher. On the other hand, non-verbal labels have a lower l^* (let us say 0.6), which leads to a lower effect of their difference on the total category difference.

For example, if the label difference, both for verbal and non-verbal labels is 0.9, this difference is multiplied with l^* for each of the labels:

Equation: $l^* * \text{diff}(L_1,L_2)$
Verbal labels: 0.9 * 0.9 = 0.81
Non-verbal labels: 0.9 * 0.6 = 0.54

As we can see, the difference of non-verbal labels is lower and consequently its contribution on category difference will be significantly lower, which will further lead to lower probability of learning the given categories.

This explanation looks satisfactory, unless we continue to question it further. There is an old saying in theoretical physics that once you cannot explain results, just put a constant in it and it will work perfectly. This saying was produced by increasing the number of constants across many equations which were used to describe fundamental laws of nature. Very few of these constants were described or explained even today.

For this reason, explicit formulation of this constant is needed. This explicit formulation is also necessary, since in it the entire nature of effects of different types of labels is encoded.

One of the explanations would be that verbal labels are (at least in adults) more reliable cues than other types of cues. From experience, adults know that labels are good category markers (Murphy, 2002). For example, more people know that Nepal is a state, but very few would recognize the flag of this state or even recognize sounds of the speech of their language (similarly with aliens in the experiments from Chapter II). Furthermore, frequency of auditory verbal labels is higher and phylogenetically older compared to written verbal labels. This property might explain the reasons why we obtained differences between auditory and verbal labels.

Finally, modality of the label can possibly influence this constant. Each modality is coded differently on the neural level. There is a possibility that auditorily presented labels are encoded with a different number of nodes compared to the ones visually presented. The same stands for the type of labels (verbal or non-verbal). This explanation is close to the number of the relevant label dimensions presented previously in this chapter.

From what is stated above, the l^* constant can be defined as following:

$$l^* = m^* * e^*$$

where m^* is the modality of the label, e^* is the experience related to the labels.

Further, the e^* constant can be specified as following:

$$e^* = f^* * r^*$$

where f^* represents frequency of occurrence of the label type, while r^* represents its reliability, meaning, how many times this type of labels proved to be reliable cue.

Again, it needs to be mentioned that this is just an attempt to quantify the obtained results. In order to fully accept this explanation, further experimental research would be needed in which each of these constants would be precisely calculated and specified.

Related to category difference, there is however one view which is thought provoking and which takes into consideration, apart from the above mentioned dimensions on which categories can differ, possible correlation of the dimensions, which can lead to overall between categories similarity. Furthermore, this model is based on Shannon's information theory. It is the model of statistical density developed by Kloos and Sloutsky (Kloos and Sloutsky, 2008).

Statistical density model

This model starts from the paradox that learning of some ill-defined categories, such as dogs and cats, is easy and effortless even for small children and without feedback, while learning some well-defined categories from the point of Boolean algebra rules is difficult even for adults who receive feedback after every learning trial. The reason for this, the model finds in so called, *statistical density* of the category.

Statistical density is defined with the following equation:

$$D = 1 - \frac{H_{within}}{H_{between}}$$

where H_{within} represents entropy within the category and $H_{between}$ entropy between two compared categories. Each of the entropies consists of two basic elements: variance of the dimensions (varying dimensions – H^{dim}) and variance of relationship between dimensions (H^{rel}). Thus, each of the entropies is calculated as the following:

$$H_{between} = H_{between}^{dim} + H_{between}^{rel}$$

and

$$H_{within} = H_{within}^{dim} + H_{within}^{rel}$$

Furthermore, each of the entropies can be formally calculated on the principles of information theory, where:

$$H_{within/between}^{dim/rel} = -\sum_{i=1}^{n} w, [\sum_{j} (p_j log_2 p_j)]$$

w_i represents attentional weight of the *i* dimension, which signifies the importance of the dimension.

Other values in this equation vary depending on which exact type of entropy is calculated. If entropy due to varying dimensions is calculated, *n* represents the number of relevant dimensions, while p_j represents probability of value *j* on dimension *i*. If entropy due to varying relations is calculated, *n* represents the number of possible dyadic correlations between dimensions and p_j represents probability of co-occurrence of the two values (*j*) on dimension *i*.

The model also presumes that mental representations of statistically dense categories are similarity-based, while representations of the statistically sparse categories are rule-based. This comes from the fact that sparse categories have one or several defining rules, which is much easier to store, than all possible dimensions and variations.

This model satisfies most of the features of a good model and it can be used to partially explain results obtained in this dissertation. It explains well category difference and easiness/difficulties to learn them. It is also a well quantified model, based on solid mathematical and informational theory background. It amends some shortcomings of the difference level model, such as correlations between features.

However, there are some *shortcomings* of this model. Primarily, the model is developed only for categorical, not continual values of dimensions (for calculation of entropy due to varying relations). Additionally, labels as an additional features and label difference level are not included in the model.

The model can be extended to be used for continual dimensions as well. In order to make it continual, when entropy due to varying relations is calculated, the following equation should be used:

$$H^{rel}_{within/between} = -\sum_{i=1}^{n} w, [\sum_{j} (|r_j| \, log_2 \, |r_j|)]$$

This equation is similar to the previous one, except that there is *r* included, where *r* represents Pearson correlation of two dimensions. Since correlation values goes from -1 to 1 (in this case from 0 to 1 for absolute value is used), correlation is a perfect numeric match with probability.

Additionally, a similar model can be developed for **label** density:

$$L_D = 1 - \frac{H_{within}}{H_{between}}$$

Where L_D represents *label density*. Label entropy could be calculated similarly to the entropy of the visual features:

$$H_{L-between} = H^{dim}_{L-between} + H^{rel}_{L-between}$$

For the label density, within-category entropy is not calculated. There are many category exemplars, but there is only one label. In that case, within category entropy is completely redundant, and it can be exchanged with the value of 1, so the entire equation becomes:

$$L_D = 1 - \frac{1}{H_{between}}$$

It would be rather challenging to calculate eventual dimensional correlation between dimensions of labels in this model, but there is always such possibility. Once these dimensions are well defined and models are developed for them, this could be the case. For this moment, we can just predict their existence.

INTERPRETATION OF CONNECTIONIST MODELLING RESULTS

As it was stated in the beginning of this section, this interpretation will be taken from the perspectives of James' hypothesis, Lupyan's hypothesis, Westermann's compound representation approach and in this dissertation developed Category learning based on the difference level model.

James' hypothesis

In Chapter V it was stated that the results obtained in the cognitive model that was developed comply with James' hypothesis (James, [1931 (1890)]). As it was stated in the Introduction chapter, James predicted that any kind of differences between labels (or anything adhered to the concepts) will drag the differences apart. Therefore, it can be concluded that higher label differences lead to easier discriminability between concepts and easier learning.

This was demonstrated in the model developed in Chapter V. In that sense, we can claim that James' hypothesis is confirmed. However, this is not the case in terms of representation: according to James' hypothesis (even if it is not explicitly stated anywhere), there are separated representations of labels and concepts. In the model from Chapter V, these representations are integrated. Additional problem which cannot fit with James' hypothesis was presumed linearity. Even though a linear function is well represented in data obtained in Chapter V, a much better approximation is a quadratic function, since small differences between labels usually do not lead to any effects on category learning.

Lupyan's language augmented thought hypothesis

In Lupyan's language augmented thought hypothesis (Lupyan et al., 2007; Lupyan, 2012a), there is a prediction of the influence of verbal labels on category learning. These effects are due to mutual activation of the two on the representational level. Lupyan presumes separate representations of verbal labels and visual features which are interconnected. Furthermore, in his connectionist model (Lupyan, 2012a), Lupyan demonstrated all these effects of verbal labels, which are separated and do influence the concepts in a feed-forward activation loop.

Concerning the model from Chapter V, Lupyan's hypothesis cannot explain the effects of verbal label differences, similarly as in behavioural results. Additionally, Lupyan presumes that there are separate representations of labels and visual features, which was not obtained in this model. For all these reasons, this hypothesis cannot be concerned as a good explanation of the results obtained in the model developed in this dissertation.

Westermann's compound representation approach

Interpretation of some portion of results obtained in Chapter V can be conducted from the perspective of Westermann's *compound representation approach* (Westermann & Mareschal, 2014; Twomey & Westermann, 2016; Capelier-Mourguy, Twomey & Westermann, 2016). This model was developed for children, but some elements can be used to explain the results from the connectionist model developed in Chapter V. Those elements are the following:

1. Verbal input is relevant for category learning and generalization,
2. Representations of visual features and verbal labels are integrated.

Verbal input is relevant for several reasons, some of which are explained in the previously mentioned Lupyan's hypotheses. Furthermore, it can be implicitly concluded (even though Westermann never explicitly stated this) that more different labels contribute to more different representations. The reason for this might be for the representations of both to be integrated (which lead us to the next element).

Integrated representations of verbal labels and visual features are very relevant for the model developed in Chapter V, since a similar modelling method was used. This has important theoretical implications, since most of the previous models considered these two types of representation separated.

From the integrated representation both labels and visual features can be extracted separately, even though these two are integrated (Westermann & Mareschal, 2014). That means that none of these two loses its identity on the representational level. This is also relevant for the model from Chapter V, even though output of this model is a single classification node.

The problem with this model is that it does not explicitly predict effects of label differences. Additionally, this model is developed for children to fit data from behavioural experiments conducted on children, but not adults. The model developed in Chapter V is exclusively developed to fit data obtained from adults.

Category learning based on the difference level

From the point of view of this approach, it is expected that label differences will be a significant factor in category learning. If the labels are more phonologically different, it is expected that categories named with them will be learned significantly faster and generalized better. This relationship is expressed with the following equation:

$$P(A,B)_T = diff(A,B)_T = diff(A,B) * l^*[diff(L_1, L_2)]$$

So the probability whether categories will be learned (discriminated) depends on the total difference between categories. This total difference is composed of the difference of visual features and difference of labels.

This hypothesis predicts that in the connectionist model, categories named with more different labels will be learned faster. This prediction was confirmed. Furthermore, the hypothesis predicts that learned categories will be generalized better, but this was not confirmed for the reasons stated in Chapter V: input was too deterministic and easy to learn. Additionally, the network was trained until it completely learned the input, which was not always the case with subjects in the experiments.

However, there are some elements which are not predicted by this hypothesis: functional dependence between the learning rate and label difference, but also the representation of labels and categories (whether is it separated or integrated). Category learning based on the difference level hypothesis predicts that the relationship between the level of difference and ease of category learning is linear. In Chapter V, we saw that this was not the case. Additionally, this hypothesis implicitly predicts that representations of labels and categories are separated. The model developed in Chapter V proved it was not necessary, since the representation level was unique for both visual features and labels.

In order to fully integrate these findings in Category learning based on the difference level hypothesis, both of these problems will be separately treated.

Functional dependency

The easiest way to integrate functional dependency obtained in Chapter V is to multiply the above stated equation with the relevant function. In order to do so, the function that was obtained in Chapter V should be calculated using variables transformed in Z values, so it will represent standardized values. After these transformations, we have the following plot of the function approximation (Figure VI-1):

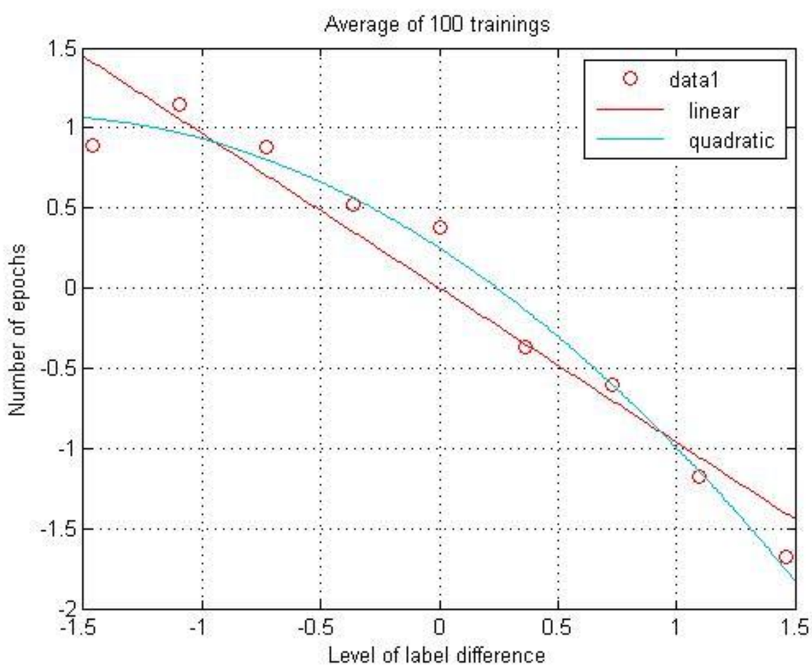

Figure VI-1: Standardized functional dependency between learning and label difference

Obtained functions are the following:

$$y = -0.96*x \text{ – linear function}$$

$$y = -0.28*x^2 - 0.96*x + 0.25 \text{ – quadratic function}$$

Obtained norms of residuals for the linear function are =0.76 and for the quadratic = 0.38.

We can see from the obtained functions and norms of residuals, that the linear function, even though its norm of residuals is higher, well explains the obtained results. Meaning, there is a negative correlation between the two variables (-.96), which is very high. On the other hand, the model explains variance for more than 90% (R^2 = .92), which is also a very high value. Since the number of training sets from which this function was approximated is quite high (N = 100), we can consider this value relevant, even though the statistical significance was not calculated.

Concerning the obtained results and property that the linear function is usually more suitable and elegant for further calculations, the linear interdependency between learning and level of label difference will be preserved. Additionally, there is possibility that the quadratic function obtained in Chapter V over-fits the data (which is less possible, since results and the data trend were consistent for a number of different training sets).

For the reasons stated above, the focal equation of Category learning based on the difference level hypothesis will not be changed to additionally fit the data. Linear dependency will be considered as relevant and its distortions will be considered as a product of error.

Representations

The second problem of the connectionist modelling data related to Category learning based on the difference level hypothesis were representations. In this hypothesis, it was implicitly presumed that mental representations of labels and visual features are separated. However, the model developed in Chapter V showed that also integrated representations are possible. For this reason, the integrated view needs to be included into this hypothesis.

Before proceeding, there must be one serious restrain stated: there is a possibility that another representational level exists between the input and hidden level. Meaning, there might exist separated representations, which get integrated on the next processing level. If that kind of network was trained in similar manner as the one from Chapter V, results would be most probably similar to the obtained ones.

The consequences of the integrated representations view are numerous, but the most relevant for this dissertation is that differences between labels and visual features are summed together, no matter from which modality they originate. For example, seeing an alien in the experiment and hearing its name, represents input which differs on a number of visual and phonological dimensions. All these dimensions are analysed and represented on one level. At this level, these differences are integrated and more easily separated and properly coded. In short, all relevant differences (no matter from which modality they come from) sum up and contribute to the overall difference and consequently to learning.

Some further consequences from this view, lead us to the conclusion that different modalities of input (visual, auditory and similar) are integrated in the same coding structure. Differences of modalities are represented to the number of input nodes, rather than to a different code. However, these kinds of questions are going out of the scope of this dissertation and will not be further examined.

3. GENERAL CONCLUSIONS OF THE DISSERTATION

The aim of this section is to give some general conclusions and state the main contributions of this dissertation.

The main contribution of this dissertation is stated as **Category learning based on the difference level hypothesis**. This hypothesis presumes that the probability of learning two categories (meaning speed of learning) depends on the overall difference level of the categories. This overall difference is composed of two parts: visual features differences and label differences. The entire hypothesis can be presented by the following focal equation:

$$P(A,B)_T = diff(A,B)_T = diff(A,B) * l^*[diff(L_1,L_2)]$$

Where $P(A,B)_T$ represents probability (easiness) that the two categories, A and B will be learned, *diff*$(A,B)_T$ – total difference between two categories, *diff*(A,B) – difference of visual features of two categories and *diff*(L_1,L_2) difference between two labels L_1 and L_2 with which the two categories A and B are labelled.

Apart from this main contribution, here will be stated some further **empirical contributions** which are mostly integrated in this hypothesis.

Results obtained in this dissertation tend to give a small illumination to the ***language and thought debate***. In the Introduction chapter it was stated that there are two schools of opinion in this debate: one which claims that thought (meaning concepts) is independent from language, while the other claims that language and thought are mutually interdependent. The results obtained in this dissertation support the latter claim, since the participants in the experiments where categories were named with more phonologically different labels, learned those categories easier and generalized better compared to those who heard phonologically less different labels. If categories were independent from labels, no effects of labels phonological difference would be recorded.

Additionally, in this dissertation it was shown that verbal and non-verbal labels do not have the same status. Only verbal labels are capable to produce facilitation in category learning, specifically verbal labels which are auditorily presented. On the other hand, non-verbal labels are not capable to produce any facilitation effects, which lead to the conclusion that ***verbal labels have special status*** compared to other types of labels.

The reason of this effect is still mysterious, but the explanation which was offered in this dissertation states that verbal labels are more complex (in terms of input), so their effects on categories are higher, since more complex input leads to higher differences between two categories and consequently to easier learning. This explanation was given along with a statistical density model (Kloos and Sloutsky, 2008) as equally relevant. Furthermore, statistical density model was amended in order to include the statistical density of the labels.

Another relevant contribution is related to ***label learning***. It was shown that it is necessary to learn labels in order to have facilitation effects on learning. This learning was induced by instruction in the experiments, where it was explicitly requested from the participants to learn labels. The results showed that the group which had the instructions, learned categories much faster compared to the one which did not have the instructions. From this it can be concluded that label learning is necessary in order to produce effects on category learning. Meaning, it is not enough for the labels to be only presented to get facilitation effects in learning.

Another important finding in this dissertation was the ***functional dependency between verbal labels and speed of category learning***, which was obtained from behavioural experimental results, especially from the connectionist model. This functional dependency is mostly linear, even though sometimes, specifically when label differences are not that high, this dependency is non-existent. Using approximation methods, the best fit was provided with quadratic function, but since linear function still explained the high level of variance, it was decided to keep linear dependency as the relevant one.

An additional contribution from this dissertation, specifically related to the cognitive model was the ***mental representations***. In the developed connectionist model, it was demonstrated that there is no need for separated mental representations of labels and visual features, since they can be integrated. This integrated representation further goes in favour to the statement that label difference influences overall category difference, which consequently leads to faster learning.

However, one restrain was stated: it is possible that there are separated representations of labels and visual features, which are integrated on the next level of processing. Furthermore, there could be another model developed which could simulate similar results. However, this does not diminish obtained results, since what was demonstrated is the relevant option, which is highly probable.

In the end, I would like to state that every scientific work is just a small drop in the ocean of the overall knowledge. What is new today becomes old tomorrow, like the drops in the oceans which slightly start sinking to the depths. However, even in the depths, this drop supports new-coming droplets which are on the top.

I hope that this dissertation presents one small drop which revealed some of the problems and make some people to ask further questions. If only one person stops and starts thinking about the stated problems and concludes that he was not thinking in the presented way or starts asking further questions, then this dissertation served its purpose completely.

R E F E R E N C E S

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly journal of experimental psychology*, *24*(2), 225-235.

Anderson, J. A., & Hinton, G. E. (Eds.). (1981). *Parallel models of associative memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Aristotel (350BC/1965), *Organon,* Kultura, Beograd

Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological review*, *105*(3), 442.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*(3), 372-400.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol.*, *56*, 149-178.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147.

Ashby, F. G., & Valentin, V. V. (2017). Multiple systems of perceptual category learning: Theory and cognitive tests. In *Handbook of categorization in cognitive science* (pp. 157-188). Elsevier.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191-208.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants?. *Journal of experimental child psychology*, *64*(1), 3-26.

Bartlett, F. C. (1932). Remembering: An experimental and social study. *Cambridge: Cambridge University*.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of experimental psychology: learning, memory, and cognition*, *11*(4), 629.

Batinić, J., Lalić, B., Taxitari, L., Ković, V. (2015). Feature position and head saliency but not labels facilitate categorisation in adults – *the 19th Conference of the European Society for Cognitive Psychology (ESCOP)* - Paphos, Cyprus

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Bloom, A. H., (1981). *The linguistic shaping of thought: A study in the impact of language on thinking in China and the West*. Hillsdale, NJ: Erlbaum.

Bloom, P., & Keil, F. C. (2001). Thinking through language. *Mind & Language*, *16*(4), 351-367.

Boas, F. (1911). Introduction: Handbook of American Indian languages (Bureau of American Ethnology, Bulletin 40, Part 1). Washington, DC: Government Printing Office.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. *Language in mind: Advances in the study of language and thought*, 61-79.

Bowerman, M. (1996). Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett (Eds.), Language and space (pp. 385–436). Cambridge, MA: MIT Press.

Boutonnet, B., Dering, B., Vinas-Guasch, N., & Thierry, G. (2013). Seeing objects through the language glass. *Journal of Cognitive Neuroscience*, *25*, 1702–1710.

Broadbent, D. E. (1958). Perception and communication. Elmsford, NY, US.

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition.*The Journal of Abnormal and Social Psychology*, *49*(3), 454.

Brown, R. W. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, *55*(1), 1.

Brown, P. (1994). The INs and ONs of Tzeltal locative expressions: The semantics of static descriptions of location. Linguistics, 32, 743–790.

Capelier-Mourguy, A., Twomey, K. E., & Westermann, G. (2016). A neurocomputational model of the effect of learned labels on infants' object representations. Cognitive Science Society.

Carroll, D. W. (2004). *Psychology of Language: 4th edition*, Thomson/Wadsworth Belmont.

Carroll, J. B. (1956). *Language, Thought, and Reality: Selected Writings of Banjamin Lee Whorf*.

Carroll, M., Von Stutterheim, C., & Nüse, R. (2004). The language and thought debate: A psycholinguistic approach. *Trends in Linguistics Studies and Monographs, 157*, 183-218.

Casasanto, D. (2008). Who’s afraid of the big bad Whorf ? Crosslinguistic differences in temporal language and thought. *Language Learning*, *58*, 63–79.

Cassirer, E. (1944). The concept of group and the theory of perception.*Philosophy and phenomenological research*, *5*(1), 1-36.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, *25*(5), 975-979.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*(1-3), 83-121.

Chomsky, N. (1957). *Syntactic Structures*. The Hague/Paris: Mouton.

Chomsky, N. (1965). Aspects of the theory of syntax Cambridge. *Multilingual Matters: MIT Press*.

Clark, H. H., & Clark, E. V. (1977). Psychology and language: An introduction to psycholinguistics. New York: Harcourt Brace Jovanovich.

Clark, E. V. (2003). Languages and representations. In D. Gentner, & S. Goldin-Meadow (Eds.), *Language in Mind - Advances in the Study of Language and Thought* -The MIT Press, 17-23.

Cohen, H., & Lefebvre, C. (Eds.). (2005). *Handbook of categorization in cognitive science*. Elsevier.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2), 240-247.

Dilthey, W., Makkreel, R. A., & Rodi, F. (1989). *Introduction to the human sciences* (Vol. 1). Princeton University Press.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93-100.

Enfield, N. J. (2015). Linguistic relativity from reference to agency. *Annual Review of Anthropology*, *44*, 207-224.

Eysenck, M. W., & Keane, M. T. (2000). *Cognitive psychology: A student's handbook*. Taylor & Francis.

Foucault, M. (2012). *The archaeology of knowledge*. Vintage.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1), 3-71.

Frege, G. (1948). Sense and reference. *The philosophical review*, *57*(3), 209-230.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, *105*(1), 218-228.

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*,*20*(3-6), 507-523.

Gentner, D., & Goldwin-Meadow, S. (2003). *Language in mind: advances in the study of language and thought*. The MIT Press.

Gentner, D., & Goldin-Meadow, S. (2003). Whither Whorf. In D. Gentner, & S. Goldin-Meadow (Eds.), *Language in Mind - Advances in the Study of Language and Thought* -The MIT Press, 3-14.

Gleitman, L., & Papafragou, A. (2005). Language and thought. *Cambridge handbook of thinking and reasoning*, 633-661.

Gleitman, L., & Papafragou, A. (2013). 32. Relations Between Language and Thought. In Reisberg, D. (Ed.). *The Oxford handbook of cognitive psychology*. Oxford University Press.

Habermas, J. (2015). *The philosophical discourse of modernity: Twelve lectures*. John Wiley & Sons.

Heider, E. R. (1972). Universals in color naming and memory. Journal of Experimental Psychology, 93, 10–20.

Hegdé, J., & Van Essen, D. C. (2003). Strategies of shape representation in macaque visual area V2. *Visual neuroscience*, *20*(03), 313-328.

Hegdé, J., & Van Essen, D. C. (2004). Temporal dynamics of shape analysis in macaque visual area V2. *Journal of neurophysiology*, *92*(5), 3030-3042.

Humboldt, W. von (1836). On language: The diversity of human language structure and its influence on the mental development of mankind (P. Heath, Trans.). Cambridge: Cambridge University Press (1988).

Ivić, I. (1987). *Čovek kao Animal Symbolicum*, Nolit, Beograd.

James, W. (1931). *The principles of psychology*, in two volumes, Henry Holt and Co, New York.

Klemfuss, N., Prinzmetal, B., & Ivry, R. B. (2012). How does language change perception: a cautionary note. *Frontiers in psychology*, *3*, 78.

Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General*, *137*(1), 52.Köhler W. (1929/1947). *Gestalt psychology*. New York: Liveright Publishing.

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological review*, *99*(1), 22.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain . *Cognition, 114(1)*, 19-28.

Ković, V., Sučević, J., & Styles, S. J. (2017). To call a cloud 'cirrus': sound symbolism in names for categories or items. *PeerJ*, *5*, e3466.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive psychology*, *58*(4), 487-524.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*(3), 265-294.

Levinson, S. C. (1996). Relativity in spatial conception and description. In J. J. Gumperz and S. C. Levinson (Eds.), Rethinking linguistic relativity (pp. 177–202). Cambridge: Cambridge University Press.

Levinson, S. C. (1997). From outer to inner space: Linguistic categories and nonlinguistic thinking. In J. Nuyts and E. Pederson (Eds.), Language and conceptualization (pp. 13–45). Cambridge: Cambridge University Press.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity* (Vol. 5). Cambridge University Press.

Locke, J. (1690/1964). A. D. Woozley (Ed.). An Essay Concerning Human Understanding. Cleveland: Meridian Books.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Lucy, J. A. (1997). Linguistic relativity. *Annual review of anthropology*, *26*(1), 291-312.

Lupyan, G. (2008). From chair to" chair": a representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*(2), 348.

Lupyan, G. (2012a). 7 What Do Words Do? Toward a Theory of Language-Augmented Thought. *Psychology of Learning and Motivation-Advances in Research and Theory*, *57*,.

Lupyan, G. (2012b). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in psychology*, *3*.

Lupyan, G. (2015). The Centrality of Language in Human Cognition. *Language Learning*.

Lupyan, G., & Casasanto, D. (2014). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(02), 167-193.

Lupyan, G., & Mirman, D. (2013). Linking language and categorization: evidence from aphasia. *cortex*, *49*(5), 1187-1194.

Lupyan, G., Mirman, D., Hamilton, R., & Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition*, *124*(1), 36-49.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077-1083.

Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*(8), 2236-2253.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive psychology*, *16*(1), 1-27.

Martinez, I. M., & Shatz, M. (1996). Linguistic influences on categorization in preschool children: A crosslinguistic study. *Journal of Child Language*, *23*(3), 529-545.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419.

McClelland,J.L., & Rogers, T.T.(2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310-322.

McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. Oxford University Press.

McWhorter, J. H. (2014). *The language hoax: Why the world looks the same in any language*. Oxford University Press, USA.

Medin, D. L. (1975). A theory of context in discrimination learning. In *Psychology of learning and motivation* (Vol. 9, pp. 263-314). Academic Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, *85*(3), 207.

Miller, N. E., & Dollard, J. (1941). Social learning and imitation.

Murphy, G. L. (2002). *The big book of concepts*. MIT press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psycological review, 92*(3), 289.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual review of Psychology*, *43*(1), 25-53.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 211.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological review*, *104*(2), 266.

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and language*, *135*, 66-72.

Piaget, J. (1926/1959). *The language and thought of the child* (Vol. 5). Psychology Press.

Piaget, J. (1936/1952). *The origins of intelligence in children* (Vol. 8, No. 5, pp. 18-1952). New York: International Universities Press.

Piaget, J., & Inhelder, B. (1950/2008). *The psychology of the child*. Basic books.

Pinker, S. (1994). The language instinct. New York: Morrow.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665-681.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological psychology*, *41*(2), 103-146.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological bulletin, 80*(1), 1.

Quillian, M. R. (1968). Semantic memory. In Minsky, M. (ed.). *Semantic Information Processing*, (pp. 227–270), MIT Press, Cambridge, Massachusetts.

Quinn P.C., Eimas P.D., Rosenkrantz S.L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception 22*, 463–475.

Regier, T, Kay, P., Gilbert, A.L., & Ivry, R.B. (2010), Language and thought: Which side are you on anyway? In B. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 165–182). Oxford, England: Oxford University Press .

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*(3), 369.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental psychology: General*, *104*(3), 192.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, *7*(4), 573-605.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (27-48) Hillsdale, New Jersey.

Rumelhart, D.E., Hinton, G.E., & McClelland, J.L. (1986), A general framework for parallel distributed processing. In D.E.Rumelhart, J.L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Volume 1, foundations.* Cambridge, MA: MIT Press.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Rumelhart, D.E., McClelland, J.L., & The PDP Research Group (Eds.) (1986). *Parallel distributed processing, Volume 1: Foundations.* Cambridge, MA: MIT Press.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*, 3-30.

Russell, B. (1903/1996). *The principles of mathematics*. WW Norton & Company.

Russell, B. (1943/2013). *History of Western Philosophy: Collectors Edition*. Routledge.

Sapir, E. (1929). The status of linguistics as a science. *Language*, 207-214.

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child development*, 309-320.

Schafer, G., Plunkett, K., & Harris, P. L. (1999). What's in a name? Lexical knowledge drives infants' visual preferences in the absence of referential input. *Developmental Science, 2(2)*, 187-194.

Slobin, D. I. (1979). The role of language in language acquisition. Invited Address, 50th Annual Meeting of the Eastern Psychological Association, Philadelphia.

Slobin, D. I. (1996). From ''thought and language'' to ''thinking for speaking.'' In J. J. Gumperz and S.C.Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge: Cambridge University Press.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner, & S. Goldin-Meadow (Eds.), *Language in Mind - Advances in the Study of Language and Thought* -The MIT Press.

Sloutsky, V. M. (2010). From perceptual categories to concepts: what develops?. *Cognitive science*, *34*(7), 1244-1286.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: a similarity-based model. *Journal of Experimental Psychology: General*, *133*(2), 166.

Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: conceptual markers or object features?. *Journal of experimental child psychology*, *111*(1), 65-86.

Sloutsky, V. M., & Lo, Y. F. (1999). How much does a shared name make things similar? Part 1. Linguistic labels and the development of similarity judgment. *Developmental psychology*, *35*(6), 1478.

Sloutsky, V. M., Lo, Y. F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child development*, *72*(6), 1695-1709.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts* (p. 89). Cambridge, MA: Harvard University Press.

Sučević, J., Savić, A. M., Popović, M. B., Styles, S. J., & Ković, V. (2015). Balloons and bavoons versus spikes and shikes: ERPs reveal shared neural processes for shape–sound-meaning congruence in words, and shape–sound congruence in pseudowords. *Brain and language*, *145*, 11-22.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, *106*(11), 4567-4570.

Thierry, G. (2016). Neurolinguistic relativity: How language flexes human perception and cognition. *Language Learning*.

Twomey, K. E., & Westermann, G. (2016). A learned label modulates object representations in 10-month-old infants. Cognitive Science Society.

Twomey, K. E., & Westermann, G. (2018). Learned Labels Shape Pre-speech Infants' Object Representations. *Infancy*, *23*(1), 61-73.

Vygotsky, L. S. (1936/1962). *Thought and language*. Cambridge, MA: MIT Press.

Wahlheim, C. N., McDaniel, M. A., & Little, J. L. (2016). Category learning strategies in younger and older adults: Rule abstraction and memorization. *Psychology and aging*, *31*(4), 346.

Watson, J.B. (1913). Psychology as the behaviorist views it. *Psychological review*, *20*(2), 158.

Watson, J. B. (1920). Is thinking merely action of language mechanisms. *British Journal of Psychology. General Section*, *11*(1), 87-104.

Waxman, S. R. (1991). Convergences between semantic and conceptual organization in the preschool years. *Perspectives on language and thought: Interrelations in development*, 107-145.

Waxman S. R., & Markow D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive psychology*,*29*(3), 257-302.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in cognitive sciences*, *13*(6), 258-263.

Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1634).

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780-7785.

Wittgenstein, L. (1922). 1984. *Tractatus logico-philosophicus*, 7-86.

Wittgenstein, L. (1953). Philosophical investigations. 1967. *Oxford: Blackwell*,*8*, 250.

Whorf, B. L. (1944). The relation of habitual thought and behavior to language.*ETC: A Review of General Semantics*, 197-215.

Whorf, B. L. (1956). Language, thought, and reality: selected writings of Banjamin Lee Whorf. (Edited by John B. Carroll).

Wolff, P., Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253-265.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223-250.

APPENDIXES

APPENDIX 1

CODDING VALUES OF LABELS USED ACROSS EXPERIMENTS

Codding	**Meaning**	**Original chapter**	**Experiment**
Silent	No label	Chapter I	Experiment 1
Min.diff	Minimal label difference	Chapter I	Experiment 1
Max.diff	Maximal label difference – also with sound symbolism incongruent with visual stimuli	Chapter I	Experiment 1
Mid.diff	Moderate label difference	Chapter I	Experiment 1
Max.diff no ss	Maximal label difference without sound symbolism	Chapter I	Experiment 2
Max.diff congr	Maximal label difference with sound symbolism congruent with visual stimuli	Chapter I	Experiment 2
nl_VIZ_MIN	Non-verbal (non-linguistic) visual label minimally different	Chapter II	Experiment 1
nl_VIZ_MAX	Non-verbal (non-linguistic) visual label maximally different	Chapter II	Experiment 1
nl_AUD_MIN	Non-verbal (non-linguistic) auditory label minimally different	Chapter II	Experiment 2
nl_AUD_MAX	Non-verbal (non-linguistic) auditory label maximally different	Chapter II	Experiment 2
ver_VIZ_MIN	Verbal visual label minimally different	Chapter III	Experiment
ver_VIZ_MAX	Verbal visual label maximally different	Chapter III	Experiment
NI_ver_AUD_MIN	Verbal auditory label without instruction (no instruction) minimally different	Chapter IV	Experiment 1
NI_ver_AUD_MAX	Verbal auditory label without instruction (no instruction) maximally different	Chapter IV	Experiment 1
NI_ver_VIZ_MIN	Verbal visual label without instruction (no instruction) minimally different	Chapter IV	Experiment 2
NI_ver_VIZ_MAX	Verbal visual label without instruction (no instruction) maximally different	Chapter IV	Experiment 2
NI_nvr_VIZ_MIN	Non-verbal visual label without instruction (no instruction) minimally different	Chapter IV	Experiment 3
NI_nvr_VIZ_MAX	Non-verbal visual label without instruction (no instruction) maximally different	Chapter IV	Experiment 3

APPENDIX 2

MATLAB SCRIPT USED FOR GENERATION OF NON-VERBAL VISUAL STIMULI IN CHAPTER II (RECEIVED FROM FROM HEDGE & VAN ESEN)

```
%Writes grayscale bitmaps to .BMP file
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%function drawGratings

        %During our actual neurophysiological recordings, the stimuli were synthesized online in real time,
        %and was customized for the preferences of neuron under study.  Among other things, the stimulus
        %size was a function of the empirically observed receptive field size of the neuron under study.
        %The present script is intended to create stimuli for off-line use.  Therefore, this script simply
        %hard-codes some stand-in values for the receptive field size.

        RFRadius=512;
  Step=1;
  stimSize=1;
  Contrast=1;
  stimRadius=512;
  Width=1;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        stimCount = 0;
  %Sinewave stimuli
  phases=pi/2;
  freqVec=[2:2:6];
  angleVec=[90:45:225];
  for(freq = freqVec);
    for(angles = angleVec);
      stim=LNG83_sine_stim(RFRadius, Step, stimSize, Contrast, freq, phases, angles, stimRadius);
                        imwrite((stim+1)/2, ['./stim_', num2str(stimCount), '.bmp']);
```

```
                              stimCount = stimCount+1;
    end
  end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%Hyperbolic Stimuli
  phaseh=0;
  angleVec=[0:22.5:67.5];
  for(freq = 1:3);
    for(angleh = angleVec);
      stim=LNG83_hyperbolic_stim(RFRadius, Step, stimSize, Contrast, freq, phaseh, angleh, stimRadius);
      imwrite((stim+1)/2, ['./stim_', num2str(stimCount), '.bmp']);
                              stimCount = stimCount+1;
    end
  end


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%Polarcc(counterclockwise); Stimuli
%First the concentric gratings
  phasep= -pi/2;
  freqr=0;
  freqc_vec=[2:5];
  for(freqc = freqc_vec);
    stim=LNG83_polarc_stim(RFRadius, Step, stimSize, Contrast, freqc, freqr, phasep, stimRadius);
    imwrite((stim+1)/2, ['./stim_', num2str(stimCount), '.bmp']);
                        stimCount = stimCount+1;
  end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%Now the radial gratings
  phasep_vec=[pi, 0, pi, 0];
  freqc=0;
  freqr_vec=[2:2:8];
  i=1;
  for(freqr = freqr_vec);
    stim=LNG83_polarc_stim(RFRadius, Step, stimSize, Contrast, freqc, freqr, phasep_vec(i), stimRadius);
```

```
  imwrite((stim+1)/2, ['./stim_', num2str(stimCount), '.bmp']);
                  stimCount = stimCount+1;
  i=i+1;
 end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
 %Now change both the concentric and the radial frequencies, *in that order*.
 freqc_vec=[1:4];
 freqr_vec=[2:2:8];
 for(freqr = freqr_vec);
  if(freqr==2)
   phasep= -pi/2;
  elseif(freqr==4)
   phasep= -pi-(pi/4);
  elseif(freqr==6)
   phasep= pi+(pi/2);
  elseif(freqr==8)
   phasep= pi-(pi/8);
  end
  for(freqc = freqc_vec);
   stim=LNG83_polarcc_stim(RFRadius, Step, stimSize, Contrast, freqc, freqr, phasep,
stimRadius);
   imwrite((stim+1)/2, ['./stim_', num2str(stimCount), '.bmp']);
                        stimCount = stimCount+1;
  end
 end

      clear all;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Generates a sinusoid.
function z = LNG83_sine_stim(rfRadius, step, stimSize, contrast, freq, phase, ori, stimRadius);

 %Check the filter val
 if(stimRadius > rfRadius);
  stimRadius=rfRadius/2;
 end
```

```
  %Create the mesh grid à la mode Matlab's meshgrid()
  num_col=length([-rfRadius:step:rfRadius]);
  basis_vec=[-rfRadius:step:rfRadius];
  rawvec=repmat([-rfRadius:step:rfRadius], 1, num_col);

  xmat=reshape(rawvec, num_col, [])';
  ymat=reshape(rawvec, num_col, []);

  %Sinusoidal foreground generating
  xvals=xmat*cos(ori/180*pi)-ymat*sin(ori/180*pi);
  yvals=xmat*sin(ori/180*pi)+ymat*cos(ori/180*pi);

  %Initalize the stimulus matrix
  z=[];
  for(i = 1:num_col);
    for(j = 1:num_col);
      rel_distance=sqrt((basis_vec(i)*basis_vec(i))+(basis_vec(j)*basis_vec(j)));
      if(rel_distance > stimRadius);
                              z(i,j)= -1;
      else
                              z(i,j)=contrast*sin((pi/stimSize*freq/rfRadius)*xvals(i,j)+phase);
      end
    end
  end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Generates a Hyperbolic Stimulus
function z = LNG83_hyperbolic_stim(rfRadius, step, stimSize, contrast, freq, phase, angleh,
stimRadius)

  %Check the filter val
  if(stimRadius > rfRadius);
    stimRadius=rfRadius/2;
  end
  %Create the mesh grid à la mode Matlab's meshgrid()
  num_col=length([-rfRadius:step:rfRadius]);
  basis_vec=[-rfRadius:step:rfRadius];
  rawvec=repmat([-rfRadius:step:rfRadius], 1, num_col);
```

```
  x=reshape(rawvec, num_col, [])';
  y=reshape(rawvec, num_col, []);

  %Initalize the stimulus matrix
  z=[];
  u=(x*cos(angleh/180*pi))-(y*sin(angleh/180*pi));
  v=(x*sin(angleh/180*pi))+(y*cos(angleh/180*pi));
  z=contrast*cos(2*pi*freq/rfRadius*sqrt(2)*sqrt(abs(u.*v))+phase);
  z=LNG83_central_filter(z, rfRadius, step, stimRadius, contrast);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%
%Generates a Polar Stimulus (CLOCKWISE);.  See LNG83_polarcc_stim(); below for COUNTER-CLOCKWISE STIMULI.
function z = LNG83_polarc_stim(rfRadius, step, stimSize, contrast, freqc, freqr, phase, stimRadius);

  %Check the filter val
  if(stimRadius > rfRadius);
    stimRadius=rfRadius/2;
  end

  %Create the mesh grid à la mode Matlab's meshgrid()
  num_col=length([-rfRadius:step:rfRadius]);
  basis_vec=[-rfRadius:step:rfRadius];
  rawvec=repmat([-rfRadius:step:rfRadius], 1, num_col);
  x=reshape(rawvec, num_col, [])';
  y=reshape(rawvec, num_col, []);

  %Initalize the stimulus matrix
  z=[];
  if(freqc>0);
    if(mod(freqr,4)==0);
      for(i = 1:(floor(2*rfRadius/step)+1));
                                             for(j = 1:(floor(2*rfRadius/step)+1));
                                                       if(x(i,j)~=0);
                                                                 radial=atan(y(i,j)/x(i,j));
                                                       else
                                                                 radial=pi/2;
```

```
                        end
                        concentric=sqrt(x(i,j)^2+y(i,j)^2);

    rel_distance=sqrt((basis_vec(i)*basis_vec(i))+(basis_vec(j)*basis_vec(j)));
                        if(rel_distance > stimRadius);
                            z(i,j)= -1;
                        else
                            z(i,j)=contrast*cos((2*pi*freqc/stimSize-
pi/2)/rfRadius*concentric+freqr*radial+phase);
                        end
                    end
                end
            else
                for(i = 1:(floor(2*rfRadius/step)+1));
                    for(j = 1:(floor(2*rfRadius/step)+1));
                        if(x(i,j)~=0);
                            radial=atan(y(i,j)/x(i,j));
                        else
                            radial=pi/2;
                        end
                        concentric=sqrt(x(i,j)^2+y(i,j)^2);

    rel_distance=sqrt((basis_vec(i)*basis_vec(i))+(basis_vec(j)*basis_vec(j)));
                        if(rel_distance > stimRadius);
                            z(i,j)= -1;
                        else
                            z(i,j)=contrast*cos((2*pi*freqc/stimSize-
pi/2)/rfRadius*concentric-pi+freqr*radial+phase);
                        end
                    end
                end
            end
        else
            if(mod(freqr,4)==0);
                for(i = 1:(floor(2*rfRadius/step)+1));
                    for(j = 1:(floor(2*rfRadius/step)+1));
                        if(x(i,j)~=0);
                            radial=atan(y(i,j)/x(i,j));
                        else
                            radial=pi/2;
```

```
                        end
                        concentric=sqrt(x(i,j)^2+y(i,j)^2);

        rel_distance=sqrt((basis_vec(i)*basis_vec(i))+(basis_vec(j)*basis_vec(j)));
                        if(rel_distance > stimRadius);
                            z(i,j)= -1;
                        else
                            z(i,j)=contrast*cos(freqr*radial+phase);
                        end
                    end
                end
            else
                for(i = 1:(floor(2*rfRadius/step)+1));
                    for(j = 1:(floor(2*rfRadius/step)+1));
                        if(x(i,j)~=0);
                            radial=atan(y(i,j)/x(i,j));
                        else
                            radial=pi/2;
                        end
                        concentric=sqrt(x(i,j)^2+y(i,j)^2);

        rel_distance=sqrt((basis_vec(i)*basis_vec(i))+(basis_vec(j)*basis_vec(j)));
                        if(rel_distance > stimRadius);
                            z(i,j)= -1;
                        else
                            z(i,j)=contrast*cos(-pi+freqr*radial+phase);
                        end
       end
      end
     end
    end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%
%Generates a Polar Stimulus (COUNTER-clockwise);.  See LNG83_polarc_stim(); above for CLOCKWISE STIMULI_
function z = LNG83_polarcc_stim(rfRadius, step, stimSize, contrast, freqc, freqr, phase, stimRadius);
```

```
%Check the filter val
if(stimRadius > rfRadius);
  stimRadius=rfRadius/2;
end

%Create the mesh grid à la mode Matlab's meshgrid()
num_col=length([-rfRadius:step:rfRadius]);
basis_vec=[-rfRadius:step:rfRadius];
rawvec=repmat([-rfRadius:step:rfRadius], 1, num_col);
x=reshape(rawvec, num_col, [])';
y=reshape(rawvec, num_col, []);

%Initalize the stimulus matrix
z=[];
if(freqc>0);
  if(mod(freqr,4)==0);
    for(i = 1:(floor(2*rfRadius/step)+1));
				for(j = 1:(floor(2*rfRadius/step)+1));
					if(x(i,j)~=0);
						radial=atan(y(i,j)/x(i,j));
					else
						radial=pi/2;
					end
					concentric=sqrt(x(i,j)^2+y(i,j)^2);

     rel_distance=sqrt((basis_vec(i)*basis_vec(i))+(basis_vec(j)*basis_vec(j)));
					if(rel_distance > stimRadius);
						z(i,j)= -1;
					else
						z(i,j)=contrast*cos((2*pi*freqc-
pi/2)/rfRadius*concentric-freqr*radial+phase);
					end
				end
			end
		else
			for(i = 1:(floor(2*rfRadius/step)+1));
				for(j = 1:(floor(2*rfRadius/step)+1));
					if(x(i,j)~=0);
						radial=atan(y(i,j)/x(i,j));
					else
```

```
                    radial=pi/2;
                end
                concentric=sqrt(x(i,j)^2+y(i,j)^2);

    rel_distance=sqrt((basis_vec(i)*basis_vec(i))+(basis_vec(j)*basis_vec(j)));
                if(rel_distance > stimRadius);
                    z(i,j)= -1;
                else
                    z(i,j)=contrast*cos((2*pi*freqc-
pi/2)/rfRadius*concentric+pi-freqr*radial+phase);
                end
            end
        end
    end
else
    if(mod(freqr,4)==0)
        for(i = 1:(floor(2*rfRadius/step)+1));
            for(j = 1:(floor(2*rfRadius/step)+1));
                if(x(i,j)~=0);
                    radial=atan(y(i,j)/x(i,j));
                else
                    radial=pi/2;
                end
                concentric=sqrt(x(i,j)^2+y(i,j)^2);

    rel_distance=sqrt((basis_vec(i)*basis_vec(i))+(basis_vec(j)*basis_vec(j)));
                if(rel_distance > stimRadius);
                    z(i,j)= -1;
                else
                    z(i,j)=contrast*cos(freqr*radial+phase);
                end
            end
        end
    else
        for(i = 1:(floor(2*rfRadius/step)+1));
            for(j = 1:(floor(2*rfRadius/step)+1));
                if(x(i,j)~=0);
                    radial=atan(y(i,j)/x(i,j));
                else
                    radial=pi/2;
```

```
                                        end
                                        concentric=sqrt(x(i,j)^2+y(i,j)^2);

    rel_distance=sqrt((basis_vec(i)*basis_vec(i))+(basis_vec(j)*basis_vec(j)));
                                        if(rel_distance > stimRadius);
                                              z(i,j)= -1;
                                        else
                                              z(i,j)=contrast*cos(-pi+freqr*radial+phase);
                                        end
      end
     end
    end
 end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%
%Convolves with a central circular filter.
function datamat = LNG83_central_filter(datamat, matRadius, mat_step, filter_rad, contrast);

 num_columns=length([-matRadius:mat_step:matRadius]);
 num_columns_in_filter=length([-filter_rad:mat_step:filter_rad, mat_step]);
 mid_point=num_columns/2;
 filter_size=filter_rad/mat_step;

 for(one_row = 1:num_columns);
  for(one_col = 1:num_columns);
   rel_coord_x=one_row-mid_point;
   rel_coord_y=one_col-mid_point;
   rel_distance=sqrt((rel_coord_x*rel_coord_x)+(rel_coord_y*rel_coord_y));
   if(rel_distance > filter_size);
                              datamat(one_row, one_col)=-1;
   end
  end
 end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%
```

APPENDIX 3

QUESTIONNAIRE USED FOR NORMING STUDY FOR NON-VERBAL AUDITORY LABELS (CHAPTER II)

Please rate the difference of the following sounds on the scale from 1 (the same) to 7 (maximal difference			
1	_1	_2	1 2 3 4 5 6 7
2	_1	_3	1 2 3 4 5 6 7
3	_1	_4	1 2 3 4 5 6 7
4	_1	_5	1 2 3 4 5 6 7
5	_1	_6	1 2 3 4 5 6 7
6	_1	_7	1 2 3 4 5 6 7
7	_1	_8	1 2 3 4 5 6 7
8	_2	_3	1 2 3 4 5 6 7
9	_2	_4	1 2 3 4 5 6 7
10	_2	_5	1 2 3 4 5 6 7
11	_2	_6	1 2 3 4 5 6 7
12	_2	_7	1 2 3 4 5 6 7
13	_2	_8	1 2 3 4 5 6 7
14	_3	_4	1 2 3 4 5 6 7
15	_3	_5	1 2 3 4 5 6 7
16	_3	_6	1 2 3 4 5 6 7
17	_3	_7	1 2 3 4 5 6 7
18	_3	_8	1 2 3 4 5 6 7
19	_4	_5	1 2 3 4 5 6 7
20	_4	_6	1 2 3 4 5 6 7
21	_4	_7	1 2 3 4 5 6 7

22	_4	_8	1 2 3 4 5 6 7
23	_5	_6	1 2 3 4 5 6 7
24	_5	_7	1 2 3 4 5 6 7
25	_5	_8	1 2 3 4 5 6 7
26	_6	_7	1 2 3 4 5 6 7
27	_6	_8	1 2 3 4 5 6 7
28	_7	_8	1 2 3 4 5 6 7

APPENDIX 4

MATLAB 2019a SCRIPT USED FOR SIMULATION OF NEURAL NETWORK (N=20) IN CHAPTER V

```
clc, clear, close all

indTrening = randperm(16);
pr_ep = [];
for d = 1:2:17
    data = load(['diff_' num2str(d) '.txt']);
    ulaz = data(:, 1:41)';
    izlaz = data(:, 42)';

    ulazTrening = [];
    izlazTrening = [];
    for i = 1:16
        r = indTrening(i);
        ulazTrening = [ulazTrening ulaz(:, 2*r-1)];
        ulazTrening = [ulazTrening ulaz(:, 2*r)];

        izlazTrening = [izlazTrening izlaz(2*r-1)];
        izlazTrening = [izlazTrening izlaz(2*r)];
    end

    ulazTest = ulaz(:, 33:end);
    izlazTest = izlaz(33:end);

    br_epoha =[];
    uspeh = [];
    N = 20;
    for j = 1:N
        net = patternnet(15);
        net.trainParam.showWindow = false;
        net.trainParam.min_grad = 0;
        net.trainParam.goal = 0;
        net.divideFcn = '';
        [net, tr] = train(net, ulazTrening, izlazTrening);
        br_epoha = [br_epoha tr.best_epoch];
```

```
        izlazPred = net(ulazTest);
        c = confusion(izlazTest, izlazPred);
        uspeh = [uspeh 1-c];
    end

    por = ['Za razliku ' num2str(d) ' i N = ' num2str(N) ' prosecan broj epoha treniranja je ' num2str(mean(br_epoha)) ', a prosecna uspesnost test skupa je ' num2str(mean(uspeh)) '.'];
    disp(por)
    pr_ep = [pr_ep mean(br_epoha)];
end

diff = 1:2:17;
figure1 = figure;
axes1 = axes('Parent',figure1);
plot(diff, pr_ep, 'o');
title(['Average of ' num2str(N) 'trainings']),
xlabel('Level of label difference')
ylabel('Number of epochs')
set(axes1,'XTick',[1 3 5 7 9 11 13 15 17]);
grid
```

APPENDIX 5

EXAMPLE OF THE CODED VALUES FOR INPUT AND OUTPUT STORED IN THE TEXT FILE USED FOR NETWORK TRAINING AND TEST (diff_7.txt, CHAPTER V)

```
1    0    1    0    1    0    1    0    0    0    0    0    0
     0    0    1    0    0    0    0    0    0    0    1    0.5
     0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5
     0.5  0.5  0.5  0.5  1
1    0    1    0    1    0    1    0    0    0    0    0    0
     0    0    1    0    0    0    0    0    0    0    1    1
     0    1    0    1    0    1    0    1    0    1    0    1
     0    1    0    1    1
1    0    1    0    1    0    1    0    0    0    0    0    1
     0    1    0    0    0    0    0    1    0    1    0    0.5
     0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5
     0.5  0.5  0.5  0.5  1
1    0    1    0    1    0    1    0    0    0    0    0    1
     0    1    0    0    0    0    0    1    0    1    0    1
     0    1    0    1    0    1    0    1    0    1    0    1
     0    1    0    1    1
1    0    1    0    1    0    1    0    0    0    0    0    0
     0    1    1    0    0    0    0    0    0    1    1    0.5
     0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5
     0.5  0.5  0.5  0.5  1
1    0    1    0    1    0    1    0    0    0    0    0    0
     0    1    1    0    0    0    0    0    0    1    1    1
     0    1    0    1    0    1    0    1    0    1    0    1
     0    1    0    1    1
1    0    1    0    1    0    1    0    0    0    0    0    1
     1    0    0    0    0    0    0    1    1    0    0    0.5
     0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5
     0.5  0.5  0.5  0.5  1
1    0    1    0    1    0    1    0    0    0    0    0    1
     1    0    0    0    0    0    0    1    1    0    0    1
     0    1    0    1    0    1    0    1    0    1    0    1
     0    1    0    1    1
1    0    1    0    1    0    1    0    0    0    0    0    0
     1    0    1    0    0    0    0    0    1    0    1    0.5
     0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5
     0.5  0.5  0.5  0.5  1
1    0    1    0    1    0    1    0    0    0    0    0    0
     1    0    1    0    0    0    0    0    1    0    1    1
     0    1    0    1    0    1    0    1    0    1    0    1
     0    1    0    1    1
```

1	0	1	0	1	0	1	0	0	0	0	0	0
	1	1	0	0	0	0	0	0	1	1	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	1							
1	0	1	0	1	0	1	0	0	0	0	0	0
	1	1	0	0	0	0	0	0	1	1	0	1
	0	1	0	1	0	1	0	1	0	1	0	1
	0	1	0	1	1							
1	0	1	0	1	0	1	0	0	0	0	0	0
	1	1	1	0	0	0	0	0	1	1	1	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	1							
1	0	1	0	1	0	1	0	0	0	0	0	0
	1	1	1	0	0	0	0	0	1	1	1	1
	0	1	0	1	0	1	0	1	0	1	0	1
	0	1	0	1	1							
1	0	1	0	1	0	1	0	0	0	0	0	1
	0	0	0	0	0	0	0	1	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	1							
1	0	1	0	1	0	1	0	0	0	0	0	1
	0	0	0	0	0	0	0	1	0	0	0	1
	0	1	0	1	0	1	0	1	0	1	0	1
	0	1	0	1	1							
1	0	1	0	1	0	1	0	0	0	0	1	0
	0	0	0	0	0	0	1	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	0	0	0	1	0
	0	0	0	0	0	0	1	0	0	0	0	1
	0	1	1	0	0	1	1	0	0	0	1	1
	0	0	0	1	0							
1	0	1	0	1	0	1	0	1	0	1	0	0
	0	0	0	1	0	1	0	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	1	0	1	0	0
	0	0	0	1	0	1	0	0	0	0	0	1
	0	1	1	0	0	1	1	0	0	0	1	1
	0	0	0	1	0							
1	0	1	0	1	0	1	0	0	0	1	1	0
	0	0	0	0	0	1	1	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	0	0	1	1	0
	0	0	0	0	0	1	1	0	0	0	0	1

	0	1	1	0	0	1	1	0	0	0	1	1
	0	0	0	1	0							
1	0	1	0	1	0	1	0	1	1	0	0	0
	0	0	0	1	1	0	0	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	1	1	0	0	0
	0	0	0	1	1	0	0	0	0	0	0	1
	0	1	1	0	0	1	1	0	0	0	1	1
	0	0	0	1	0							
1	0	1	0	1	0	1	0	0	1	0	1	0
	0	0	0	0	1	0	1	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	0	1	0	1	0
	0	0	0	0	1	0	1	0	0	0	0	1
	0	1	1	0	0	1	1	0	0	0	1	1
	0	0	0	1	0							
1	0	1	0	1	0	1	0	0	1	1	0	0
	0	0	0	0	1	1	0	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	0	1	1	0	0
	0	0	0	0	1	1	0	0	0	0	0	1
	0	1	1	0	0	1	1	0	0	0	1	1
	0	0	0	1	0							
1	0	1	0	1	0	1	0	0	1	1	1	0
	0	0	0	0	1	1	1	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	0	1	1	1	0
	0	0	0	0	1	1	1	0	0	0	0	1
	0	1	1	0	0	1	1	0	0	0	1	1
	0	0	0	1	0							
1	0	1	0	1	0	1	0	1	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	1	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	1
	0	1	1	0	0	1	1	0	0	0	1	1
	0	0	0	1	0							
1	0	1	0	1	0	1	0	0	0	0	0	1
	0	0	1	0	0	0	0	1	0	0	1	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	1							

1	0	1	0	1	0	1	0	0	0	0	0	0
	0	1	0	0	0	0	0	0	0	1	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	1							
1	0	1	0	1	0	1	0	0	0	0	0	1
	0	1	1	0	0	0	0	1	0	1	1	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	1							
1	0	1	0	1	0	1	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	1	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	1							
1	0	1	0	1	0	1	0	1	0	0	1	0
	0	0	0	1	0	0	1	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	0	0	1	0	0
	0	0	0	0	0	1	0	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	1	0	1	1	0
	0	0	0	1	0	1	1	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							
1	0	1	0	1	0	1	0	0	1	0	0	0
	0	0	0	0	1	0	0	0	0	0	0	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5	0							

B I O G R A P H Y

BIOGRAPHY

Bojan Lalić was born on 10th of November, 1979 in the city of Peć, where he finished his primary and secondary schooling. He studied psychology at the University of Priština, Faculty of Philosophy, at which, as the student of the final year, he was awarded the title of the best student at the Faculty of Philosophy (diploma "Eminent student"). He graduated in 2012 (graduate psychologist) with the graduation thesis "Symbolic and Analogue Coding Systems of Mental Representations in Long-Term Memory" earning the highest mark (10 – A+).

In 2012 he started his master studies (research master) at the University of Belgrade, Faculty of Philosophy, within the Psychology department. During the first year of the studies, he completed all exams with the highest marks (10 on average – A+) and defended his master thesis, titled: "Processing of visually and auditory presented inflective nouns in Serbian language" in 2014.

At the end of 2013, he started his PhD studies at the University Belgrade, Faculty of Philosophy within the Psychology department. During his studies, he spent two months on a study visit at the Max Plank Institute for Psycholinguistics in Nijmegen, Netherlands (2014). Additionally, he spent a further two months conducting research in the United States of America, at the University of Wisconsin Madison, in Lupyan's lab at the Psychology department (2016).

He was awarded with a stipend from the "Borislav Lorenc" Foundation in 2016, for his study visit in the United States.

He has published several scientific papers, of which two are published in journals from SCI list (M22 and M23) and has presented papers at many conferences worldwide.

Образац 5.

Изјава о ауторству

Име и презиме аутора: Бојан Лалић

Број индекса: 4П13/0008

Изјављујем

да је докторска дисертација под насловом:

The role of label features and label remembering in concept formation: behavioural, neural and cognitive modelling approach
(Улога својстава и памћења именитеља на формирање појмова: бихејвиорални, неурални и приступ когнитивног моделовања)

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 09.12.2019.

____

Образац 6.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора: Бојан Лалић

Број индекса: 4П13/0008

Студијски програм: Психологија

Наслов рада: *The role of label features and label remembering in concept formation: behavioural, neural and cognitive modelling approach*
(Улога својстава и памћења именитеља на формирање појмова: бихејвиорални, неурални и приступ когнитивног моделовања)

Ментор: проф. др Вања Ковић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао ради похрањена у **Дигиталном репозиторијуму Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, 09.12.2019.

____

Образац 7.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

The role of label features and label remembering in concept formation: behavioural, neural and cognitive modelling approach
(Улога својстава и памћења именитеља на формирање појмова: бихејвиорални, неурални и приступ когнитивног моделовања)

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)

2. Ауторство – некомерцијално (CC BY-NC)

3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)

4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)

5. Ауторство – без прерада (CC BY-ND)

6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 09.12.2019.

____

1. **Ауторство**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. **Ауторство – некомерцијално**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. **Ауторство – некомерцијално – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. **Ауторство – некомерцијално – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. **Ауторство – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. **Ауторство – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.