

Title of Dissertation/Titolo della Dissertazione

Language and Thought: the Case of Labels and Categorisation in Infants

Author/Autore

Floris Mara

(Year/Anno)

2020

Examination Committee/Commissione di esame:

Prof. Peter Brössel

Prof. Vincenzo Crupi

Prof. Kim Plunkett

The copyright of this Dissertation rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

End User Agreement

This work is licensed under a Creative Commons Attribution-Non-Commercial-No-Derivatives 4.0

International License: <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

You are free to share, to copy, distribute and transmit the work under the following conditions:

- *Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).*
- *Non-Commercial: You may not use this work for commercial purposes.*
- *No Derivative Works - You may not alter, transform, or build upon this work, without proper citation and acknowledgement of the source.*



In case the dissertation would have found to infringe the polity of plagiarism it will be immediately expunged from the site of FINO Doctoral Consortium

Contents

Introduction	1
1 Assessing the role of labels	10
1.1 Labels and other auditory inputs	10
1.1.1 Experiments with a Silence control condition	11
1.1.2 Experiments comparing sounds	14
1.1.3 Experiments with ecologically plausible sounds	15
1.1.4 Experiments with language	18
1.1.5 Comparing different kinds of words	19
1.1.6 Advancements during development	22
1.1.7 Conclusion	25
1.2 How labels shape categories	26
1.2.1 The experiments in Plunkett et al. (2008)	27
1.2.2 The experiments in Althaus & Westermann (2016)	30
1.2.3 Interpretation of the effects	31
1.2.4 Filling the perceptual space	35
1.2.5 Other studies	37

1.2.6	Conclusion	39
2	Top-down theories	41
2.1	Natural Pedagogy	43
2.1.1	Infants are sensitive to ostensive signals	43
2.1.2	Natural Pedagogy and the effects of labels in categorisation	49
2.1.3	Conclusion	50
2.2	The Referential Role of Labels	51
2.2.1	The argument in Waxman & Gelman (2009)	53
2.2.2	Critical assessment	58
2.2.3	Labels highlight commonalities	60
2.2.4	Conclusion	63
2.3	The Label Feedback Hypothesis	64
2.3.1	The paradox of the effects of language	64
2.3.2	The on-line effects of labelling	66
2.3.3	Implications for the studies on infants	69
2.3.4	Conclusion	70
3	Bottom-up theories	71
3.1	Labels as features in Sloutsky's studies	73
3.1.1	The first experiments	73
3.1.2	Further research: labels and inferences	78
3.1.3	Sloutsky's arguments and critiques	79
3.1.4	Conclusion	82

Contents	iii
3.2 Labels as features in Plunkett et al. (2008)	83
3.2.1 Interpretation of the results	84
3.2.2 Plunkett's argument	87
3.3 Neural networks as explanations	91
3.3.1 Computationalism	91
3.3.2 Gliozzi et al.'s (2009) self-organising map	93
3.3.3 Psychological and biological plausibility of the model	101
3.3.4 Predictions made by the model	104
3.3.5 Predictions and accommodations	109
3.3.6 Conclusion	112
4 General Conclusion	113
4.1 Overview of the advancements	113
4.1.1 The role of labels	113
4.1.2 Top-down theories	115
4.1.3 Bottom-up theories	118
4.2 Discussion of the theoretical options	121
4.2.1 Top-down and bottom-up	121
4.2.2 Labels and features	123
4.2.3 The model	125
4.2.4 Final remarks	126
References	128

List of Figures

1.1	The developmental funnel of acoustic stimuli that facilitate categorisation. Notice that some stimuli are spontaneously accepted, others need habituation. Human language at first is enough, regardless of the presence of a labelling expression. After the first year of life the presence of a labelling expression is fundamental.	24
1.2	The stimuli used in the familiarisation phase by Plunkett et al. (2008).	27
1.3	The test stimuli used by Plunkett et al. (2008).	28
1.4	Table of the experiments in Plunkett et al. (2008).	28
1.5	The stimuli used in the familiarisation phase by Althaus & Westermann (2016).	30
1.6	The stimuli used in the test phase by Althaus & Westermann (2016).	31

1.7	A possible graphical representation of the results of Plunkett, Hu & Cohen (2008). Experiment 3, left, and Experiment 5, right. The stimuli are the same; the different labelling pattern changes the perceived categories.	32
1.8	A possible graphical representation of the results of Althaus and Westermann (2016). When the objects are presented in silence, they are considered as a single category, left. When the objects are paired with two different labels, they are split into two categories, right.	33
1.9	The stimuli in Althaus & Westermann (2016).	35
1.10	Some of the stimuli used in Landau & Shipley (2001).	36
1.11	The stimuli used in Fulkerson & Waxman (2007).	39
2.1	An example of the stimuli typically used Waxman and colleagues, Fulkerson & Waxman (2007).	52
2.2	The stimuli in the upper part are an example of those used in the familiarisation phase; those in the lower part are those used in the test phase (Althaus & Mareschal, 2014).	61
2.3	An example of the stimuli used by Althaus & Plunkett (2016).	62
2.4	An example of the coloured patches used by Winawer et al. (2007).	65
2.5	The stimuli used by Lupyan et al. (2007).	69
3.1	(Sloutsky & Lo, 1999, p. 1484)	73
3.2	(Sloutsky & Lo, 1999, p. 1484)	75

3.3	Overview of the experiments in (Plunkett et al., 2008)	84
3.4	(Plunkett et al., 2008)	89
3.5	(Plunkett et al., 2008)	90
3.6	The representation of the network provided by Gliozzi et al. (2009).	94
3.7	Duta 2018.	95
3.8	Duta 2018.	96
3.9	The table of results reported in (Gliozzi et al., 2009, p. 724). .	99
3.10	Mather & Plunkett (2011)	107
3.11	Gliozzi et al. (2009).	111
4.1	The stimuli of the Narrow Condition in Plunkett et al. (2008).	125

Introduction

In the past decades, attention has risen around the topic of language and categorisation: can the way we label objects affect the way we categorise them? In this thesis, I will deal with this question in reference to the literature about infants.

Proving that labels can affect categorisation is not an isolated issue, it can be framed in at least two long-standing philosophical debates: Linguistic Relativity and the Cognitive Penetrability of Perception.

The Sapir-Whorf Hypothesis

According to the Sapir-Whorf Hypothesis, also called Linguistic Relativity, the language we speak influences the way we think, at least, this is what states the most popular version of this theory.

The exact content of this theory is unclear; the name *Sapir-Whorf Hypothesis* itself is improper, as Edward Sapir and Benjamin Lee Whorf never proposed it as a co-authored theory, although Sapir was Whorf's mentor¹. It is doubtless that Sapir had a significant influence on Whorf's work, but it is also undeniable that Whorf independently developed the hypothesis, this is

¹It is also unclear who is responsible for the diffusion of the name *Sapir-Whorf Hypothesis*. Some claim that it is due to the linguists Eric Lenneberg and Roger Brown, some others claim that the linguist Harry Hoijer mentioned it in a paper. The high diffusion it had is due to the psychologist John Carroll (Koerner, 1992).

the reason why it is often referred as the *Whorfian* hypothesis.

Whorf's most famous fragment states that:

“We dissect nature along lines laid down by our native language. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscope flux of impressions which has to be organized by our minds - and this means largely by the linguistic systems of our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way - an agreement that holds throughout our speech community and is codified in the patterns of our language. The agreement is of course, an implicit and unstated one, but its terms are absolutely obligatory; we cannot talk at all except by subscribing to the organization and classification of data that the agreement decrees. We are thus introduced to a new principle of Relativity, which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated.” (Whorf, 1940)

Linguistic Relativity, initially, has been accepted as an undeniable fact and psychologists and sociologists studied it as an axiom; only in the Seventies, the increasing interest for psychological universalism cast some doubt on its validity. It has then been the rejected, especially in its stronger version which is *linguistic determinism*.

Linguistic determinism is the theory that states that *everything* we can think is determined by the language we speak. It is impossible to identify who proposed this version of Linguistic Relativity, as it can not be inferred from Whorf's work and no one ever claimed its authorship. Anyhow, it is rather implausible that language determines all our cognitive activity, and therefore it is not surprising that it was rejected.

However, in the Nineties, a new interest for Linguistic Relativity rose, and the empirical investigations on the topic become kept growing, as witnessed by the publication of some volumes about it (Gumperz & Levinson, 1991; Niemeier & Dirven, 2000; Verspoor & Putz, 2000).

Recent developments

Nowadays, weaker versions of the hypothesis are still being tested, and some theoretical advancements were made. The existing studies tackle different areas where effects of Linguistic Relativity could be found; in particular, it is important to define what part of language affects what cognitive aspect.

A possible hypothesis is that language could affect reasoning, for example, counterfactual thinking could depend on the use of the subjunctive, which is not an element of every human language². However, most of the existing studies focus on whether language can impact perception, conceptualisation and categorisation, which, as we will see, are not the same thing. Among the most common topics there are studies on colours (e.g., Franklin et al., 2008; Regier et al., 2007; Winawer et al., 2007), on the effects of grammatical gender (e.g., Cubelli et al., 2011), on object perception (e.g., Malt et al., 1999) and motion perception (e.g., Athanasopoulos et al., 2015).

The above-mentioned studies are different for topics and methods, but they all identify some common aspects of language on thought:

1. The effects of language are *on-line*; namely, they are active as long as the language is being used. It means that these effects disappear when participants of an experiment are given a verbal interference task ³.

²This hypothesis turned out not to be true (Au, 1983; Bloom, 1981; Liu, 1985).

³See section 2.3.

2. Some effects are active as long as verbalisation is required; for example, during an experiment, some effects may be available only if the participant knows that she will have to give a verbal answer.
3. The effects of language on thought are not rigid. Language does not permanently and deeply affect cognition, as Linguistic determinism claims. The effects of language create *habits* rather than rigid schemes.
4. Bilingual speakers can switch from *habitual schemes* when they change the language.

Finally, a helpful distinction was drawn by Lucy (1997), who identified three different levels at which language can influence cognition:

1. Having a language, any language, may affect thought in comparison with animals or pre-verbal infants.
2. Speaking a specific language could make a difference; for instance, English or Italian could affect cognition in different ways.
3. Inside the same language, there could be differences depending on the linguistic abilities of the speakers.

The cases described by Whorf, and most of the studies, usually address the second option, but the case study that I will analyse in this thesis belongs to the first and the third options. This is possible because the participants of the studies I examine are mainly prelinguistic infants. The experiments, or at least some of them, compare the effects of labels on categorisation with categorisation in silence, and they also test whether having different labels shapes categories.

Cognitive Penetrability of Perception

The debate on Linguistic Relativity intersects another debate, the one on the Cognitive Penetrability of Perception (CPP); the thesis of CPP is that perceptual experience can be influenced by our beliefs, desires or mental states. It is a controversial and debated thesis both on the theoretical and on the empirical side on many different levels. On the theoretical side, for instance, CPP would have a crucial fallout on epistemology: if higher levels of cognition impact perception, its role as a “truth-preserving source of knowledge of the world” is not guaranteed (Vetter & Newen, 2014).

Those who claim cognitive impenetrability think that perception is a module (e.g., Carruthers, 2006; Fodor, 1983, 2000; Sperber & Wilson, 2002) and that its processes are encapsulated, which preserves their role as a source of reliable knowledge. On the contrary, those who claim that perception can be penetrated also deny the existence of modules in a strict sense and accept that knowledge is grounded on perception, even if there is not a truth-preserving perception mechanism.

The debate on what does it mean that perception is penetrated led to a fine-grained description of what is perception, what is cognition and where is the boundary between the two; the most recent studies even started questioning the existence of such a boundary (see Beck, 2018; Burnston, 2017; Montemayor & Haladjian, 2017; Vetter & Newen, 2014). The core part of this debate is focused on whether early vision can be affected by higher processes; even if perception as a whole is discussed, most of the studies focus only on vision.

For the purposes of this thesis, it is important to keep in mind the existence of these two debates because they both help in framing the effect of labels on categorisation.

Linguistic Relativity and CPP have an intersection: Linguistic Relativity holds that language can affect *any* level of cognition; CPP holds that perception is affected by higher levels of cognition. The common subset is the one where language affects perception. The experiments I will describe in this thesis belong to this intersection, they investigate whether only one specific aspect of language – naming – can affect a specific cognitive process – categorisation.

Definitions

Before discussing the effects of labels on categorisation it is fundamental to define what are labels and, in particular, what are categories. Concepts and categories are often used as synonyms, especially by psychologists, but it is worth to disambiguate their use.

Labels and names

The core issue I will deal with in this thesis is whether the effects on categorisation stem from top-down processes because labels refer, or if they can impact categorisation also in a bottom-up manner because can count as additional perceptual features.

Following Plunkett et al. (2008), in this thesis I will rarely use the term *names* because names refer and it may be the case the infants do not consider labels as referent yet. *Label* is a neutral term which does not imply any commitment on its role. Labels will be called names when there is evidence that they are used in a referential way.

Concepts

The notion of *concept* is as pervasive in cognitive science as it is unclear. Machery (2009) described the currently available definitions of concepts and

claimed that cognitive scientists should abandon the very notion of concepts and replace it with the terms which refer to what he calls the fundamental kind of concepts: *prototype*, *exemplar* and *theory*. He defines concepts as:

“Within cognitive science, a concept of x is a body of information about x that is stored in long-term memory and that is used by default in the processes underlying most, if not all, higher cognitive competences when they result in judgments about x .”(Machery, 2010, pp. 195-196).

Machery (2010) also describes the properties of concepts:

- Concepts can be about classes of objects (such as CAT)⁴, events (such as RUNNING), substances (such as GOLD) and individuals (such as IMMANUEL KANT).
- Concepts can be used in multiple processing: they can be used for categorisation, induction, linguistic comprehension and others.
- Concepts can vary over time and are different across individuals.
- Concepts are used *by default* by cognitive processes⁵.

The experiments I will describe are mainly conducted with infants, the notion of concepts that is needed to describe their behaviour is minimal. What infants are required to do is to look at sets of images, or plastic toys, which are more or less similar, and then with an experimental procedure called *novelty preference task* it is assessed whether they consider some new items as familiar or not. Depending on their preferences, it is possible to infer whether they formed one or more categories.

⁴Small caps are conventionally used to indicate concepts.

⁵“By default” here means that it is preferentially available and that it spontaneously comes to the mind.

There is no need to posit any form of representational content for these objects because there is no information stored other than their physical appearance. This is the reason why I am reluctant to claim that labels impact concepts, even if psychologists often state it. The effects of labels on categorisation may be the basis of concept learning, but this is a question which goes beyond the purpose of this thesis.

Categories

Although *concepts* and *categories* are often interchangeable, in this thesis I will try to keep them separate for the above-mentioned reason: when I claim that infants can categorise objects I do not want to commit to the fact that they possess concepts for those objects, even if it is possible. In particular, categorisation as a process is not identical to have categories. A minimal definition of *categorisation* need to account for the fact that even infants are able to sort objects into classes, without possessing any information other than their physical appearance. Furthermore, the categories they form do not need to be stable over time.

Categorisation, which I will consider just as the ability to sort the objects into classes, is an essential process both for animals and human beings.

To understand the categorisation as it is intended in the experiments described in this thesis, another useful distinction is the one between *conceptual categories* and *perceptual categories* (see Mandler, 2007). In the experiments infants are exposed only to the physical features of the objects, therefore, the only way to cluster them in categories is on the basis of their physical similarity. They are tasks of object identification and not conceptual understanding.

It is plausible to think that infants can learn perceptual categories well before they display conceptual abilities, at least because this faculty is shared

with other animals(Mandler, 2007). This does not mean that infants do not possess conceptual abilities at all, what I am claiming is that it is not necessary to postulate any conceptual understanding to explain the results of the experiments discussed in this thesis.

Overview of the thesis

The first chapter concerns the role of labels in a strict sense. I will first review the existing studies to show for the idea that the effects in categorisation actually depend on labels and not on their being sounds or language. Secondly, I will describe two effects, a grouping effect and a segregation effect. The second chapter reviews the theories which claim that labels act in a top-down manner. The third chapter, finally, will review the theories which claim that labels, instead, act as bottom-up stimuli, with particular attention to the use of neural networks as part of the explanations of these theories.

Chapter 1

Assessing the role of labels

1.1 Labels and other auditory inputs

When dealing with the role of labels in categorisation, the first step is assessing whether the effects of language, if any, actually depend on labels and not on a general auditory input. In this section, I will argue that the effects on categorisation initially depend on a broad variety of auditory stimuli that becomes increasingly narrow during development. By the second year of life, in fact, only count nouns affect the categorisation process. First, I will analyse the studies in which there is a comparison between the categorisation process in silence and the very same process (with the same stimuli) in the presence of a verbal label. Then I will consider the studies in which there is a comparison between the effects of sounds and those of non-labelling expressions. Finally, I will focus on the studies that highlight the specific role of count nouns as compared to language in general and adjectives. None of the studies conducted so far includes a direct and systematic comparison of all these variables.

The purpose of this section is to show that consideration of the existing literature and a comparison of the studies supports the claim that labels

do play a role in categorisation. I take into account 17 studies, published between 1995 and 2016, largely uniform with respect to their research design. Most of the experiments on this topic are eye-tracking studies with infants (3 to 26 months) using a novelty preference procedure; only two of them had a different research design.

The novelty preference task relies on the principle that infants show a preference for novelty. Usually, in the first phase of the experiments, infants are familiarised with a set of visual stimuli (such as drawings of animals) all belonging to the same category. Items are presented one at the time, they are either drawings shown on a screen or plastic toys. In the second phase, infants are tested with two new objects, one belonging to the familiarised category and one completely novel. If the participant shows a preference for the novel object, this is taken as a sign that the other object is considered similar to the familiarised examples, and it is meant to belong to the same category. If, on the contrary, the participants prefer the within-category object or shows no preference, it is inferred that the category presented in the familiarisation phase was not learnt. The experiments vary in the type of auditory stimuli presented in the familiarisation phase (e.g., tones, sentences, novel nouns) along with the visual items, and in the age groups that have been tested.

1.1.1 Experiments with a Silence control condition

The first result to assess is whether the presence of a label in the familiarisation phase makes a difference in the categorisation process when compared to a condition in which the stimuli were presented in silence. Only 3 out of the 17 studies considered in this section had a Silence control condition: most of them just compared the presence of a labelling expression to the one of a non-labelling expression or a sound. In order to state that labels have an

advantage in categorisation, however, including a Silence condition is crucial.

For example, in some studies, there seems to be an advantage of labels over sounds, but no Silence baseline is employed. In the absence of such a condition, the claim that labels enhance categorisation remains unsupported. It could be that sounds hinder categorisation (so-called “overshadowing effect”, Best et al. (2011)) and that labels do not have any effect. The advantage of labels would then be only apparent and seem to be at work just because the comparison with categorisation in silence is lacking. The only three studies that overcome this problem are Plunkett et al. (2008), Althaus & Mareschal (2014) and Althaus & Westermann (2016).¹

In Plunkett et al. (2008) two sets of stimuli were presented in the familiarisation phase, a Broad and a Narrow Condition. All the stimuli were sketched animals that varied in the size of the neck, legs, tail, and ears. In the Broad Condition, the four features combined randomly, whereas in the Narrow Condition they were correlated (e.g., long neck with short legs and vice-versa) in order to form two clusters of stimuli. If the stimuli were presented in silence, the Broad Condition would lead to the formation of one single category and the Narrow Condition to the formation of two categories. The Narrow Condition yielded a binary categorisation again if paired with two consistent labels, while if the labels were randomly assigned, it was not possible to measure any proof of categorisation with the novelty preference task. Finally, if the Narrow Condition was paired with a single label, the

¹Actually, there are other studies (Balaban & Waxman 1997; Haaf et al. 2003) in which some of the stimuli presented in the familiarisation phase are presented in silence, rather than with a sound or a linguistic expression, but in these experiments the stimuli presented in silence are not tested with a separate novelty preference task. The only available finding is that there is a quick decrease of attention for the stimuli presented in silence when compared to those presented with language. This lack of attention was usually considered enough to conclude that labels do have an effect as compared to silence, but they should have tested the categorisation process of the item presented in silence rather than accepting just the decreasing of attention in the familiarization phase as significant.

stimuli were then considered as belonging to the same category. Althaus & Westermann (2016) used a similar research design: their set of stimuli consisted of drawings of invented animals, and it was possible to segregate them in two visual categories in much the same way as in the Narrow Condition used by Plunkett and colleagues. When the stimuli were presented in silence or with a single label, in the test phase, the overall average stimulus was considered familiar, and only one category was formed. When the stimuli were presented with two consistent labels, the two sub-category prototypes were considered familiar, and two categories were formed. When the stimuli were presented with two consistent sounds (a tingling bell and a xylophone tone sequence), then it was not possible to measure any preference at testing.

The studies just mentioned were both conducted on 10-month-old infants, whereas the experiments of Althaus & Mareschal (2014) concerned a group of 8-month-olds and one of 12-month-olds in four conditions: Silence, Labelling expression, Non-labelling expression, Sound. With the first group, it was not possible to measure any categorisation of the visual stimuli presented. With the second group, instead, categorisation was achieved both with a labelling and a non-labelling expression, but not in the absence of any auditory stimuli or with a non-linguistic sound.

These three experiments show that categorisation occurs at least sometimes even in silence, but that labels can disrupt categories that would be formed otherwise, or enable categorisation not taking place in silence. Appropriately, at least some of the above-mentioned experiments (Althaus & Westermann 2016; Althaus & Mareschal 2014) did have comparisons between a Silence condition, a Sound condition and a Label condition. These cases show that the effect of labels is not only apparent, as it would be if sounds hindered categorisation.

1.1.2 Experiments comparing sounds

The number of experiments that included a condition in which a sound was compared to language is more substantial. Also, the variety of sounds used in these experiments is quite broad. Eight out of 17 studies in this section compared sounds to labelling and non-labelling expressions. It is crucial to prove that the facilitative effect on categorisation does depend on language (or labels) and not merely on the presence of any auditory input. In principle, any sound could help in focusing attention, thereby leading to a positive outcome in categorisation.

Balaban & Waxman (1997) had familiarised 9-month-olds with a set of visual stimuli paired either with a tone (a 400 Hz sine wave tone) or with a noun phrase (“A pig!” or “A rabbit!”). The proportion of infants looking at the novel object at test was higher for those in the Word condition than for those in the Tone condition.

Haaf et al. (2003) tested two groups of infants, 9-month-olds and 15-month-olds. Each of the two groups was in turn split into two conditions: basic-level and superordinate-level. In the familiarisation phase, they were exposed to some visual stimuli (20 plastic toys, animals or vehicles) accompanied by a labelling phrase (“Look at the toma/bicket”), a five-note melody or non-labelling repetitive mouth sounds. The data suggest that labelling phrases facilitated global categorisation, but not basic-level categorisation (that was always achieved), over non-labelling sounds both at 9 and 15 months of age. There is also a sensitivity to the source of the auditory stimuli, and it undergoes some changes as infants grow up: 9-month-olds accomplish categorisation at global level, despite the source of the auditory input; 15-month-olds achieved global categorisation only when the experimenter directly uttered labelling phrases. According to the authors, the fact that basic-level categorisation was achieved despite the presence of an

auditory stimulus may depend on the low perceptual variability among the stimuli: a higher perceptual similarity among stimuli makes the category easier to detect.

Similar results were reported by Fulkerson & Waxman (2007). They tested a group of 6-month-olds and a group of 12-month-olds with a set of figures depicting dinosaurs. The auditory stimuli were presented to half of the infants accompanied by a naming phrase (“Oh look, it’s a toma/modi” or “Do you see the toma/modi?”) and to the other half with two sequences of pure tones (400 and 800 Hz). Naming phrases were uttered by a female voice in the infant-directed speech register and recorded for presentation; the tone sequences were created to match the naming phrases in timing, duration and volume. In the test phase, 12-month-olds in the Word condition demonstrated a reliable novelty preference, whereas those in the Tone condition performed at chance level; 6-month-olds showed the same effect.

A more recent study (Ferry et al., 2013) used the same set of stimuli as Fulkerson & Waxman (2007), but with a group of 3/4-month-olds. Their results were similar to those of the previous study: labels do have a facilitative effect on categorisation that one does not achieve with tones. Finally, both Althaus & Mareschal (2014) and Althaus & Westermann (2016) had a Sound condition after which it was not possible to measure any preference in the test phase. All the experiments considered here point in the same direction: sounds do not improve infants’ performances in categorisation tasks.

1.1.3 Experiments with ecologically plausible sounds

The sounds used in the experiments discussed above were mainly tones. In this section, I will discuss some experiments in which other kinds of sounds were used. The reason why I keep them separate is that the complexity of this last group of sounds may make them ecologically more plausible. It may

appear unsurprising that pure tones fail to affect categorisation, for they are usually not employed as communicative signals. Even if it is established that infants are able to detect their native language when they are born (J. Werker J.F. Gervain, 2013), there might still be many variations in the kind of signal that affects their categorisation process. An ecologically plausible sound may thus be necessary to impact categorisation.

In a study already mentioned, Balaban & Waxman (1997) tested a group of 12/13-month-olds and one of 9-month-olds. They had a Tones condition, a Words condition, and a Content-filtered words condition. The content-filtered words were obtained by filtering the original, computer-digitised, phrases with an electronic filter system in order to remove high frequencies. These stimuli were recorded on tape for presentation and were matched in loudness to the other word phrases and tone sequence. Balaban & Waxman (1997) found that during the test phase the preference for the novel object was stronger for those who heard proper words; content-filtered words enhanced the preference for the novel item only if compared to tones, their effect was not as strong as words.

An interesting result was found by Hespos and Waxman (2013): they provided evidence for the idea that infants up to 4 months may accept non-verbal sounds as communicative signals. The set of stimuli they used is the same as Fulkerson & Waxman (2007), their participants were divided into three groups: 3-month-olds, 4-month-olds and 6-month-olds. The three groups were tested in two different conditions: lemur vocalisations and backward speech. The lemur vocalisations were chosen because, although they differ from human vocalisations, they still share certain acoustic properties with infant-directed speech. The data from this trial are similar to those obtained with human speech: there is a categorisation effect for 3/4-month-olds, but the 6-month-olds did not perform above chance. Hespos and Waxman hy-

pothesise that what matters may be the complexity of the sound employed. In order to test this hypothesis, they adapted the experiment with backward speech, which is as complex as normal speech. In this experiment, it was not possible to measure any evidence of categorisation. There was no measurable cognitive advantage of backward speech, while there was an effect of primate vocalisation.

Perszyk et al. (2016) go further in exploring the effects of primate vocalisations. Their experimental design is the same as Hespos and Waxman (2013), except that immediately before the familiarisation phase infants (4-month-olds) were exposed to lemur vocalisations embedded in a 10-min soundtrack of classical music (a Mozart piano concert). During the exposure phase, lemur vocalisations were part of the infants' acoustic environment. The effect of the exposure manipulation was rather strong: infants had a robust preference for the novel image, more than in the same experiment without the exposure. This effect was not observed with the exposure to backward speech, as tested in the second experiment. Finally, in the third experiment, Perszyk and Waxman increased the time of the exposure from a few minutes before the test up to 6 weeks. Infants, at home, listened to the soundtrack every day in the first week, every two days day in the second week, and three times per week thereafter. Afterwards, infants (who were by now 6 months of age) took part in the experiment. Unlike infants in the first experiment, they did not listen to the soundtrack upon arrival at the laboratory. Instead, they were directly engaged in the object categorisation task. Even if they had not heard the lemur vocalisations for a mean of two days, they still performed as well as in the first experiment.

This study suggests that the nature of the sounds that are accepted as communicative signals depends on what kind of sounds is present in the environment. If the outcome of the analysis of the experiments which tested

the effect of sounds was not promising (no effect was measured), the same cannot be said of complex sounds and in particular of primate's vocalisations: they do seem to have a weak effect on categorisation. Furthermore, this effect can be reinforced via habituation.

1.1.4 Experiments with language

It could be the case that what makes a difference is not the presence of a noun: a general linguistic auditory input could be enough. In other words, as claimed by the advocates of the so-called "Natural Pedagogy" view (Csibra Gergely 2011, 2009), being involved in a communicatively rich environment might be enough to improve categorisation performance. Sharing attention, through language, may improve the learning process by directing saliency. In an experiment by Waxman & Markow (1995), a group of thirty-two infants (9,3 to 20,1 months, mean age 13 months) was presented with a set of forty lightweight plastic toys. Infants were randomly assigned to the Noun or No Word condition.

In the Noun condition, in the familiarisation phase, the experimenter addressed the infant directly by saying: "[Infant's name]! Look a(n) X". In the No Word condition, the attention was caught just by calling the infant by name, without the labelling expression. Participants in the No Word condition showed no decrease of attention during the familiarisation phase, nor a preference for an object in the novelty preference task. The effect was measurable only for those in the Noun condition. The results are fairly robust, but the age group is rather broad, therefore some differences in performance across different age groups of participants might have remained unnoticed. Other studies show evidence for the fact that labels, and not merely language, improve categorisation. In Fulkerson & Haaf (2006) a group of 12-month-olds was familiarised with a set of novel visual stimuli in two conditions: with a

labelling expression (“Look, a mot/fep!”) and with a non-labelling expression (“Look, here’s one!”). Then they were tested in a word extension task: two objects were simultaneously presented, one belonging to the familiarised category and the other completely novel; participants were then asked to pick up the one that belonged to the familiarised category. Only the items that were paired with a labelling expression were recognised.

Another study that investigates the role of labelling expressions vs non-labelling expressions is again by Althaus & Mareschal (2014). They tested a group of 8-month-olds and one of 12-months-old. For the first group, the results were the same regardless of the presence of the label. In the older infants, instead, only if the objects were paired with a label, infants looked at the common features of the items. Common features are those that determine the membership to a certain category. This experiment proves that even if the results of the novelty preference task were the same, at 12 months infants are sensitive to the influence of labels, and they show it in the pattern of eye movements during the familiarisation.

1.1.5 Comparing different kinds of words

Once it is established that a sound, in general, is not sufficient to affect categorisation, and that language by itself is not enough, one still has to consider which kind of words are accepted. Two different issues arise:

1. possibly what matters is the presence of a noun in general, regardless of the consistency of the link between a particular object and a particular noun;
2. perhaps the word does not have to be a noun: an adjective may have the same effect.

A study by Waxman & Braun (2005) addresses the first point: they

are interested in whether applying a consistent name to a set of distinct objects is crucial to categorisation, or whether variable names might serve the same function. They tested a group of 12-month-olds infants with the same procedure of Waxman & Markow (1995), but with an additional Variable Noun condition. In the Consistent Noun condition, the same name was always paired with a specific set of objects (e.g., four different animals), while in the Variable Noun condition the noun varied for every object of the same set. Infants that were in the Consistent Noun condition obtained almost the same score of those in the 1995 experiment, whereas those in the Variable Noun condition provided no evidence of categorisation, as in the No Word condition.

These findings suggest the idea that a string of speech not embedding a noun is not enough, and that even if there is a noun, it has to be consistent in order to induce a facilitative effect. One of the experiments of Plunkett et al. (2008) reaches the same conclusion. If the labels are inconsistent, it is not possible to measure any preference in the test phase.

Concerning the second question – the nature of the words that promote categorisation – some experiments by Waxman and colleagues have compared the effects of nouns and adjectives. Waxman & Markow (1995) tested 12/13-month-olds both in the Novel Noun condition (“Look, and X!”) and in the Novel Adjective condition (“Look, the X-ish one!”). No differences in categorisation were measured with the novelty preference task, but it is essential to notice that the standard deviation of the age of their participants was very high: they ranged from 9.3 to 20.1 months.

Waxman (1999) tested her hypothesis about the specific role of count nouns with a group of 13-month-olds. The set of stimuli used included 40 small plastic toys, and subsets of toys were arranged in order to promote categorisation either on the basis of shape or on the basis of colour. In the

test phase, the two objects belonged to the same basic-level category, but they had different properties (e.g., colour/texture), or the other way round. The authors wanted to test whether a new noun would lead to the recognition of the category (e.g., horse, carrot) or to the recognition of the property (e.g., colour, texture). The findings can be summarised into two main points:

- Novel words (both nouns and adjectives) draw attention to objects: to average looking time in the familiarisation phase was shorter in the No Noun condition. Since there was spoken language on every trial, the attentional effect must be attributed to the novel noun/adjective and not to language in general
- By 13 months infants begin to be sensitive to the distinction between count nouns and adjectives: infants performed better at novelty preference task if they were in the Novel Adjective condition, although they performed better in the Novel Noun condition compared to the No Word condition.

Although the results indicate a difference between adjectives and count nouns, it is not clear at all if infants are sensitive to the role of adjectives or if they just discriminate count nouns from all other words. It is important to remember that at 13 months infants start to produce and comprehend count nouns, while they learn names for colours only later. A similar result comes from Waxman & Booth (2001). They conducted two experiments in which they tested the ability to construe a property-based category or a basic-level category, starting from the very same set of objects. In all conditions, infants were familiarised with the same set of objects that shared both category membership and salient object property. The study showed that infants are sensitive to the difference between nouns and adjectives only if they are older than 14 months.

A significant issue is how do infants distinguish a name from an adjective: it probably depends on the grammatical position of the word in the sentence. A study investigating the role of labels in early induction in 16-month-olds reaches a similar conclusion (Keates & Graham, 2008). The experiments at issue are quite different from those discussed so far: instead of measuring the familiarity with a visual stimulus, Keates and Graham tested the willingness to infer hidden properties of objects. Infants relied on labels only when they were presented as count nouns, namely embedded in a sentence with a grammatical marker for a count noun, such as an article (e.g., "This is a blick"). The same results did not arise with a non-labelling sentence, if the label was marked as an adjective (-ish), or if the name was presented alone. Concerning this last condition, it is worth noting that in other experiments a label presented alone did lead to some effect on categorisation. The experiments of Plunkett et al. (2008), as reported in Hu's doctoral thesis (Hu, 2008), had the nonce word presented without a grammatical marker after a carrier word (e.g., "Look! Dax/Rif"). Even if the results of Keates and Graham go in the opposite direction, it could be argued that a single word is taken as a count noun. Infants usually expect new words to be nouns (see Mc Donough et al., 2011).

1.1.6 Advancements during development

In the first part of this section, I tried to explain why the facilitative effect of categorisation depends on count nouns, rather than on generic auditory stimuli, on language as such, or on words of any kind. But the experiments discussed also enable us to draw a developmental trajectory in the kind of signals that facilitate categorisation (see Ferguson & Waxman, 2016, for a similar suggestion).

1. I examined three experiments with infants aged 3/4 months: in one of

them there was categorisation with labelling expressions, but not with tones (Ferry et al., 2013), the other two proved that there is categorisation in the presence of human language and primate vocalisations, but not with backward speech (Hespos and Waxman 2013; Perszyk Waxman 2016).

2. Concerning 6-month-olds, there is no effect of tones (Fulkerson & Haaf, 2006), unless the tones are used in a communicatively rich environment as signals (Ferguson & Waxman, 2016). A similar result has been obtained with primate vocalisations: they do not exert any effect on categorisation (Graham et al., 2013), unless there is a long habituation period (Perszyk et al., 2016).
3. At 9/10 months labels have a facilitative effect on categorisation, either embedded in a sentence or uttered alone. Sounds do not affect categorisation. (Balaban Waxman 1997; Fulkerson Haaf 2003; Plunkett, Hu Cohen 2008; Althaus Westerman 2016).
4. Similar results have been found in 12/13-month-olds (Fulkerson Waxman 2006, Althaus Mareschal 2014). But, additionally, in this age group infants are sensitive to the difference between nouns and adjectives (Waxman, 1999). This is true also of 14-month-olds (Waxman & Booth, 2001), furthermore, they are sensitive to the consistency of names (Waxman & Braun, 2005). In 13-month-olds a flexibility similar of the one described for 6-month-olds has been found by Woodward & Hoyne (1999), infants at that age may still accept a sound-object link in a communicative context.
5. Older infants show a perceptual narrowing of the signals that they accept while categorising new objects, only count nouns have an effect. The tolerance for a sound-object link found by Woodward & Hoyne

(1999) cannot be observed any more at 20 months. Interestingly, Namy & Waxman (1998) discovered that at 18 months even a gesture could facilitate categorisation, but it doesn't work anymore at 26 months.

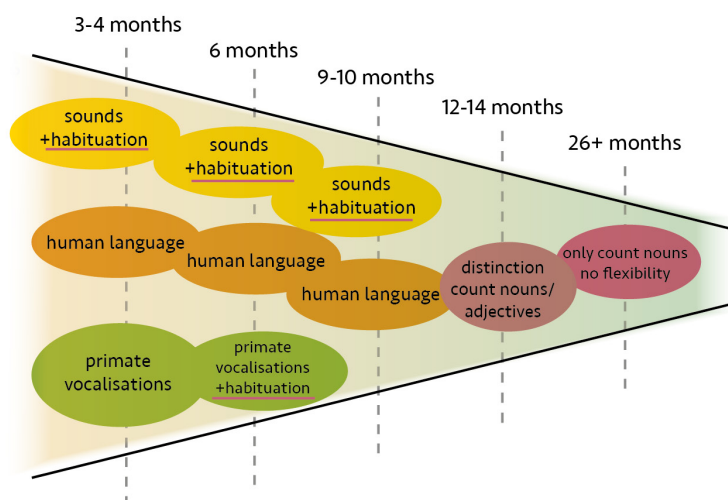


Figure 1.1. The developmental funnel of acoustic stimuli that facilitate categorisation. Notice that some stimuli are spontaneously accepted, others need habituation. Human language at first is enough, regardless of the presence of a labelling expression. After the first year of life the presence of a labelling expression is fundamental.

Taken all together these findings indicate a perceptual narrowing in infancy of the kind of signals that facilitate categorisation. In a first phase, newborns accept a quite broad set of stimuli, the presence of a certain kind of stimulus in the environment makes it eligible as a communicative one. The same results are found at 6 months, but only after habituation, otherwise only human language is accepted. At 10 months only language has an effect, other sounds are no longer accepted. When they turn 14 months, in order to influence categorisation, there must be a noun, adjectives are no longer accepted. By the second year of life only count nouns produce facilitative effects on categorisation: the range of accepted stimuli becomes more and more narrow.

1.1.7 Conclusion

In the first part of this section, I tried to show that, even if more research is needed, facilitative effects on categorisation do seem to depend on labels, at least after 14 months of age. This result is compatible with the experiments conducted with adults and children in which the relevance of labelling expressions is less controversial (e.g., Lupyan, 2012b). It is harder to claim that with infants younger than 14 months labels have an influence. At an early stage, newborns are sensitive both to human language and to primates' vocalisations, later on they show a clear preference for human language. What is clear is that ecologically implausible sounds do not facilitate categorisation.

What about the reasons causing this developmental tunnel? Not much can be said yet. One can only speculate that at first every communicative signal helps in sharing attention between the infant and the experimenter and that attention improves learning. In this respect, supporters of Natural Pedagogy may be right. As the infant grows, shared attention is no longer enough, otherwise it would not be possible to explain why non-labelling expressions, used to catch attention, do not have any effect. When infants start speaking, and in particular when they are able to distinguish between nouns and adjectives, it is evident that a noun is needed to obtain some effect on categorisation. Children know that count nouns refer to objects.

The perceptual narrowing suggested here deserves further investigation. In particular, more experiments with a Silence condition remain crucial in order to have a proper control condition for the way categorisation is affected. As for the purpose of this thesis, the present literature is enough to support the idea that labels have a role in categorisation in young infants.

1.2 How labels shape categories

In the previous section, I reviewed the evidence in favour of the idea that labels, at least at a certain age, have some effects on categorisation. In this section, I will describe what the nature of this effect is.

Many psychologists claim that labels “facilitate” categorisation, but it is not clear what this statement means. Most of the experiments that led to this idea are conducted with the novelty preference procedure: in a first phase infants are familiarised with a set of visual stimuli all belonging to the same category, and in the test phase they see two different images, one of them is a member of the familiarised category, the other one is new. If children look longer at the new object, it is considered as a sign that infants recognise the other object as “familiar”. The same experiment is usually repeated by pairing the familiarised stimuli with a sound or a label. As we have seen in the previous section, it should be repeated in silence too.

The fact that this kind of experiments is usually run without a Silence control condition weakens the claims based on the data obtained, but there is another reason to doubt of most of the experiments in this field, and it has been pointed out very clearly by Plunkett et al. (2008). Infants are usually familiarised with a single category and a single label, for instance, they can be familiarised with a set of dinosaurs presented with a label (“Look, a toma”) or a sound. In the test phase, they see both a fish and a new dinosaur; if they look longer at the fish, it means that they recognise the dinosaur as familiar and the fish as new. What usually happens is that there is a novelty preference only when the stimuli are presented with a label, whereas in silence, the same preference is not found.

This experimental set-up does not tell much about categorisation and the role of labels, and it could be the case that the presence of the label had

just the function of tuning attention. If there is not a direct comparison of the same familiarised stimuli with and without a label, but there is only a comparison between labels and sounds (or consistent and inconsistent labels, nouns and adjectives, etc.), it is not possible to estimate what labels do in comparison with the same condition without auditory inputs; therefore if labels have any effect, we do not know what it is.

For this reason, there are only two studies which can be considered reliable in defining the role of labels on categorisation: Plunkett et al. (2008) and Althaus & Westermann (2016). Only in these two studies it is possible to assess what happens in the absence of any auditory stimulus and with a label.

1.2.1 The experiments in Plunkett et al. (2008)

The set of stimuli used by Plunkett et al. (2008) consisted of drawings representing sketchy animals, Fig.1.2.



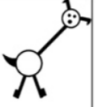

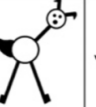
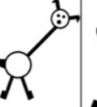






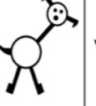
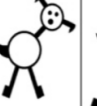

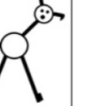
Broad Condition	 1155	 1515	 2244	 2424	 4422	 4242	 5511	 5151
Narrow Condition	 1122	 1212	 2211	 2121	 4455	 4545	 5544	 5454

Figure 1.2. The stimuli used in the familiarisation phase by Plunkett et al. (2008).

All the animals share the shape of the body and the face (two circles) and have different legs, necks, ears and tails. The legs and neck vary in their lengths, the tail can assume different degrees of thickness, and the distance between the ears can vary as well. Each of the four features can have five possible values (numbered from 1 to 5); therefore, every animal corresponds

to the combination of four numbers.²

They have two different experimental conditions: the Broad Condition and the Narrow Condition. In the Broad Condition, the combination of the features has no restriction, whereas in the narrow Condition values on one dimension are correlated to values of the other one (e.g., long necks were always paired with short legs and vice-versa). The stimuli used in the test phase, which immediately follows the familiarisation, are new animals: the overall prototype (3333), the extreme exemplar with low values (1111) and the one with high values (5555), see Fig.1.3.

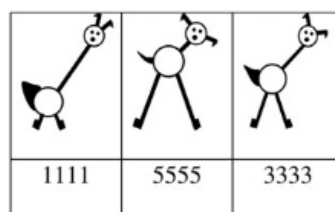


Figure 1.3. The test stimuli used by Plunkett et al. (2008).

They conducted 5 experiments, as reported in Fig.1.4.

Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
Broad Condition	Narrow Condition	Narrow Condition	Narrow Condition	Narrow Condition
Silence	Silence	2 consistent labels	2 random labels	1 label
1 category	2 categories	2 categories	No categorization	1 category

Figure 1.4. Table of the experiments in Plunkett et al. (2008).

In the first experiment, children are familiarised with the Broad Condition in silence, at the test they preferred the objects with the extreme values (1111/5555) rather than the overall mean object (3333), which is the

²There is not a correlation between the number and the dimension, the features represented by number “1” not necessarily are short legs or short neck.

prototype of the category. The fact that the two extreme exemplars are recognized, and the prototype accumulated less looking time during the test, is considered as a proof that the objects presented during the familiarization phase are recognised as members of the same category because the objects which received a longer looking time are perceived as new.

In the Narrow Condition the pattern of the preferential looking time is the opposite: after being exposed to the stimuli of the Narrow Condition, infants look longer at the prototype (3333) rather than the two extreme exemplars (1111/5555). Each of the two extreme exemplars, from the point of view of geometrically measured similarity, was close to one of the two categories of the Narrow Condition and, as it was to be expected, it was not the preferred object at test. The prototype, instead, was perceived as more interesting. This finding means that two categories were recognised, but the perceptual space between them was not filled. These first two experiments were crucial to assessing that the manipulation of the stimuli could lead to the recognition of one or two categories depending on the familiarised stimuli.

The third experiment was identical to the second, but the stimuli were paired with two consistent labels: one of the two subcategories of the Narrow Condition was paired with the label “rif”, the other one with the label “dax”. The presence of these additional auditory stimuli did not change the outcome of the novelty preference task: the two subcategories of the Narrow Condition remained unchanged.

In the fourth experiment, instead, the stimuli were pseudo-randomly paired with the same two labels of the third experiment, which means that there was no correlation between the two subcategories and the two labels. This mismatch had a disruptive effect on the previously formed categories; their choice at test was not different from chance.

Experiment 5 is the most interesting in the perspective of describing the

effects of labels on visual categorisation: when the stimuli of the Narrow Condition were accompanied by one single label, the results were similar to those of the first experiment where the familiarised stimuli were those of the Broad Condition. The use of one label yielded to only one category, and it means that calling a set of perceptually dissimilar objects (which in silence are considered as two categories) with the same name can lead to the formation of only one category. This effect, which I will call “grouping effect” broadly corresponds to the naïve idea that if different objects are called by the same name, they are considered as members of the same category.

1.2.2 The experiments in Althaus & Westermann (2016)

The opposite effect can be found in a more recent study by Althaus & Westermann (2016). They created a set of stimuli which in silence is considered as one category that can be split into two categories if the stimuli are consistently paired with two labels. Their stimuli were drawings of animals created with a morphing software: the items of the set were images created by mixing two different animals along a continuum. The features of these animals were not individually manipulated, but there was a holistic difference between each animal. As for the Narrow Category used by Plunkett and colleagues, the stimuli could be divided into two subcategories; this was obtained by leaving a gap in the middle of the morphed stimuli.

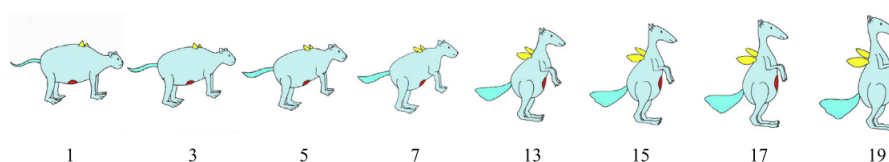


Figure 1.5. The stimuli used in the familiarisation phase by Althaus & Westermann (2016).

Their test is more complex than the one used by Plunkett and colleagues:

in order to make sure that a familiarity effect could be excluded, they compare the overall prototype with the two subcategory prototypes, and then they compare them with a novel stimulus, Fig.1.6.

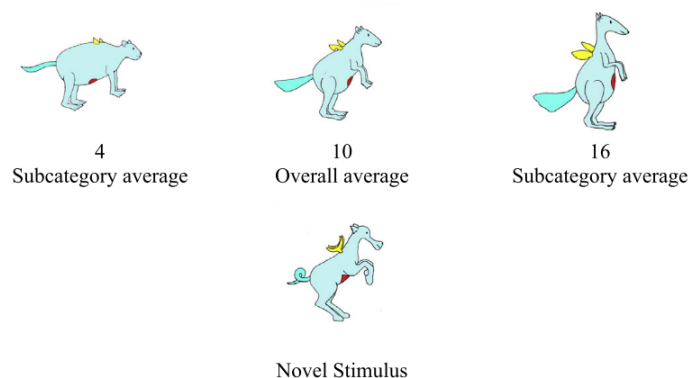


Figure 1.6. The stimuli used in the test phase by Althaus & Westermann (2016).

When the objects are presented in silence, in the first experiment, they are considered as a single category: the overall average stimulus is perceived as familiar. The same result is achieved when the same label accompanies all the stimuli. Instead, when two different labels accompany the two subcategories, the overall average stimulus accumulates a longer looking time, which means that infants split the category in two. Hearing a sound does not have any effect on categorisation.

The effect found here is the opposite of the one previously found by Plunkett et al. (2008) in the following aspect: a specific set of stimuli in silence is considered as a single category and using two different labels modifies this outcome, and the stimuli are then considered as two categories.

1.2.3 Interpretation of the effects

Before discussing these effects, the differences between the two studies and comparing them to the existing literature, I will briefly discuss the limits of

the novelty preference task and the notion of category which is used hereafter. The novelty preference procedure is based on the idea that infants look longer at new items rather than to familiar ones; Fantz (1964) is the first one who described this effect. Despite being used quite broadly in Psychology, the habituation paradigm is still unclear, for instance, infants sometimes show a familiarity preference if habituation is not long enough, younger infants often express a familiarity preference, especially if the stimuli are complex Cohen (2004). However, even if we assume that in the two studies mentioned above the novelty preference was not an experimental artefact and it was reliable, this does not eliminate all the problems linked to this procedure. This method does not offer a direct measure of categorisation, but only an indirect measure: it only shows when a particular stimulus is perceived as more novel when it is compared to another one. The two effects listed above could be graphically described as in Fig. 1.7.

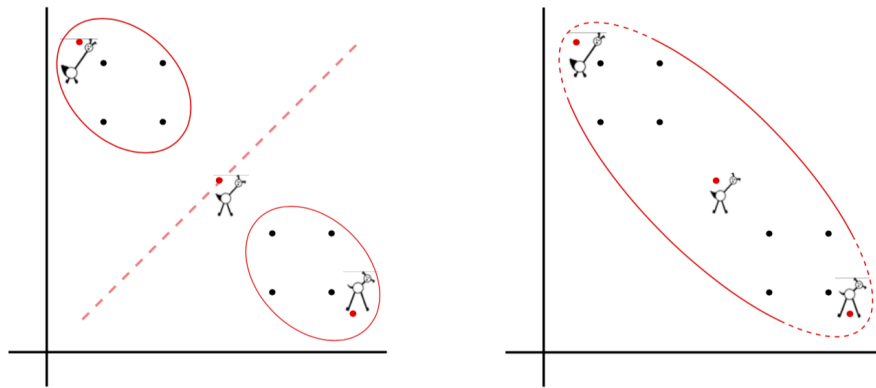


Figure 1.7. A possible graphical representation of the results of Plunkett, Hu Cohen (2008). Experiment 3, left, and Experiment 5, right. The stimuli are the same; the different labelling pattern changes the perceived categories.

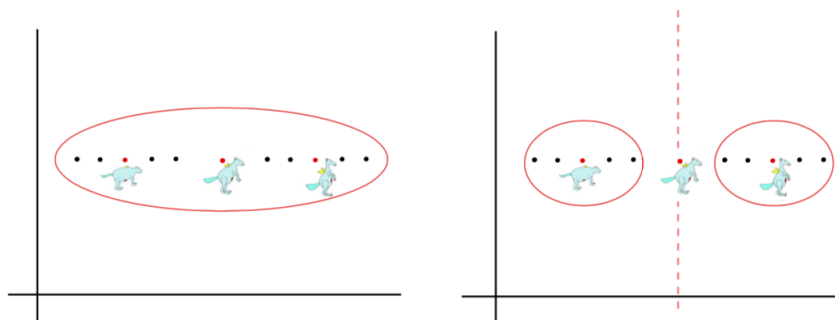


Figure 1.8. A possible graphical representation of the results of Althaus and Westermann (2016). When the objects are presented in silence, they are considered as a single category, left. When the objects are paired with two different labels, they are split into two categories, right.

It is worth noticing that in Plunkett’s experiments, during the test, the two stimuli used to test the two possible subcategories were novel extreme exemplars. In contrast, in Althaus’ experiments, they were the prototypes of the two subcategories.

If presented in silence, Plunkett’s stimuli are split into two categories: the extreme stimuli are perceived as more familiar than the prototype, which accumulated a longer looking time. It means that there is a gap between the two categories. The presence of a label fills the gap between the two subcategories: the prototypical item is recognised as familiar, and plausibly all the items between it and the two categories are recognised. The data do not allow to say whether the two extreme exemplars are considered as belonging to the category or not. In the novelty preference task, two objects are compared, the one that receives a longer looking time is just more novel than the other one, but the unchosen object could be either familiar or not.

That is to say that the novelty preference procedure gives the relative preference between two objects, any judgement about the inclusion or exclusion in a category is inferred indirectly. If an object which geometrically is in the middle of the category is perceived as more novel than the two extremes,

it is legitimate to say that there is a gap, but if the object is extreme, if it is in a peripheral area of the category, it could be either included or excluded in the category when it is preferred to the prototype. It is not possible to say what happens in the boundaries of the categories given the data.

For the very same reason, when in the silence condition of Althaus' experiments the overall average prototype is considered more familiar than the two subcategory prototypes it just means that there is no gap among them, it is plausible to think that the two subcategory prototypes are included in the category. What the label does, in this case, is making unfamiliar the overall average item and, presumably, other items in the middle.

This reverse pattern deserves more attention as it opens some interesting future research questions. The Narrow Condition in Plunkett's experiments is specifically designed to obtain two categories. Plunkett and colleagues reproduced the same stimuli used by Younger (1985); Younger invented them to test whether infants could exploit correlation of attributes to create new categories. Their original purpose was to test the so-called "Correlated attribute hypothesis" according to which natural categories are not arbitrary, but they carve-up the world according to clusters of features (Medin & Smith, 1984; Rosch et al., 1976). Younger and colleagues discovered that infants could actually exploit correlations among attributes, or, at least, they found that the stimuli they created were naturally divided into two groups. What is surprising is that the stimuli used by Althaus & Westermann (2016) are not *naturally* segregated into two categories. If we look at the way the stimuli are created, we can notice that, as they morphed two images, they had of the entire continuum of the stimuli and that they deliberately decided to leave a big gap in the middle. They took only one every two morphed animals, and they discarded the five animals in the middle.

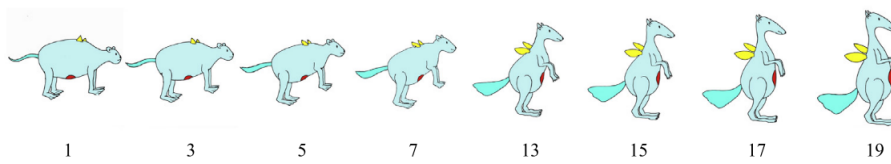


Figure 1.9. The stimuli in Althaus & Westermann (2016).

This big gap between the two groups of stimuli is meant to guarantee two categories, but it is not the case. The stimuli, in silence, are considered as a single category. Whatever are the principles that control the way infants categorise, Plunkett’s Narrow Condition is split into two categories, and Althaus’ stimuli are not, even if they were meant to be so. It would be interesting to investigate this kind of phenomena further.

The other interesting future development of these data concerns the nature of the novelty preference procedure itself and the role of prototypes. In Althaus’ experiments, in silence, the overall prototype was considered more familiar than the two subcategory prototypes. This result is surprising because the two subcategory prototypes are closer to the exemplars seen during the familiarisation phase and the overall prototype, instead, is relatively far from them. This experiment seems to show that prototypical effects on categorisation are so strong that even if infants do not see items close to the prototype, not only they recognise it as familiar, but it is also considered as *more familiar* than two items closer to the already seen exemplars. More research is needed to shed light on these questions: what drives natural categorisation? How could an un-seen prototype be more familiar than seen exemplars?

1.2.4 Filling the perceptual space

Concerning the grouping effect, there is a study which seems to show a similar result despite the methodological differences. Landau & Shipley (2001) tested three groups (2-year-olds, 3-years-olds and adults) with the same set

of stimuli consisting of two standard objects and six objects created by morphing the initial two along a continuum; they were randomly assigned to the Same Label or Different Labels condition, Fig.1.10.

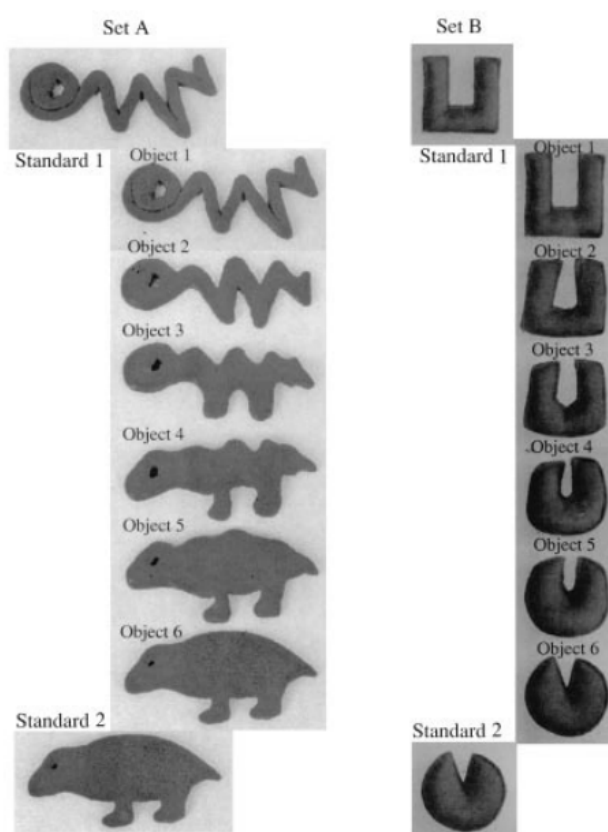


Figure 1.10. Some of the stimuli used in Landau & Shipley (2001).

Participants first observed two standard objects either called by the same name (“This is a blicket”) or with two different names (“This is a blicket/steps). They were then exposed to the other stimuli, those in the Same Label condition were asked if the object was a blicket and those in the Different Label condition were asked if it was a blicket or a step. In the Same Label condition, participants were likely to call all the objects by the same name. In the Different Label condition, the name was generalized only to the most similar exemplars. Even if the age groups and the procedure are different compared

to Plunkett's and Althaus' studies, this study shows that when two objects are given the same name, all the intermediate objects, with respect to perceptual similarity, are considered members of the same category. These findings corroborate the idea that calling perceptually different objects by the same name yields to the filling of the perceptual gap between them.

1.2.5 Other studies

The reason why only the two already mentioned studies have been considered is that they are the only studies in which it is clear that a label can modify the shape of a category that can be formed in silence. In the existing literature, this requirement has not been fulfilled. It is true though that there is a substantial body of empirical research on this topic, as mentioned in the previous section.

The lack of a Silence control condition makes the evaluation of the previous research at least uncertain; nonetheless, the analysis in the previous section on the kind of auditory stimuli that may affect categorisation lead to a positive conclusion. Thanks to the comparison of the existing studies, it is possible to say that if there is an effect of labels (more specifically count nouns) on categorisation, this effect depends on their being names, not on their being sound, or linguistic sounds. This finding sheds light on the experiments in which there was a comparison between a Label condition and a No-Label condition. If it accepted that only labels have an effect, at least after a certain age, the experiments in which there is a condition in which stimuli were paired with a sentence, but without a label, could be considered significant.

Only a few experiments fulfil these requirements: they have to be conducted with infants older than 10 months because before that age the facil-

itative effects could depend on language broadly conceived, and there must be a comparison between a condition with a label and a condition without a label. Experiments comparing sound and words can not be considered because of a possible overshadowing effect of sounds (Fulkerson & Haaf, 2006; Haaf et al., 2003; Waxman & Markow, 1995). Even if these experiments are acceptable from the point of view of the comparison among different kinds of auditory stimuli, their experimental set-up is not adequate to describe what is the exact effect on categorisation. Their results seem to prove that some forms of categorisation are not available without a label: it is not possible to measure any preference at test in the condition without a label, the category is recognised only when a label is present. What is surprising is that during the test one of the two items always belongs to a completely novel category (e.g., infants are familiarised with a set of dinosaurs and at the two items of the test are a new dinosaur and a fish). This result, compared with the existing literature is quite hard to interpret; the interpretation according to which the category “dinosaur” was not recognised without a label is weak. It is plausible to think that it is recognised even in silence, but for some reason, the fish is preferred at test only when there is a label.

Another possibility is that the stimuli used by Waxman and colleagues induced some sort of “superordinate-level” categorisation, and the stimuli used by Plunkett et al. (2008) and Althaus & Westermann (2016) kept categorisation at a “basic-level”. The perceptual variability in the latter studies is relatively low. It is not implausible to think about them as items belonging to the very same basic category. In Waxman’s studies, instead, the animals used during the familiarisation phase are quite different from each other, Fig.4.1.1.











Trial	Word	Tone	Left Screen	Right Screen
Familiarization 1	Look at the toma! Do you see the toma?	---- -- -- -- -- -- -- --		
Familiarization 2	Look at the toma! Do you see the toma?	---- -- -- -- -- -- -- --		
Familiarization 3	Look at the toma! Do you see the toma?	---- -- -- -- -- -- -- --		
Familiarization 4	Look at the toma! Do you see the toma?	---- -- -- -- -- -- -- --		
Familiarization 5	Look at the toma! Do you see the toma?	---- -- -- -- -- -- -- --		
Familiarization 6	Look at the toma! Do you see the toma?	---- -- -- -- -- -- -- --		
Familiarization 7	Look at the toma! Do you see the toma?	---- -- -- -- -- -- -- --		
Familiarization 8	Look at the toma! Do you see the toma?	---- -- -- -- -- -- -- --		
Test				

Figure 1.11. The stimuli used in Fulkerson & Waxman (2007).

Maybe the perceptual variability among the dinosaurs is so high that there is a novelty preference only when a label highlights their belonging to the same category. Basic level categories, instead, are easier to detect. The novelty preference procedure only tells what item is perceived as more interesting; it does not mean that the no categorisation occurred.

1.2.6 Conclusion

The existing studies indicate that it is possible to describe a “grouping effect” and a “segregation effect” at least in some circumstances. The role of labels, so far, is to increase or decrease the perceived similarity. Therefore, if some items share the same name, they may be considered as belonging to

the same category. If some other items without any auditory stimulus look quite similar, but they have different names, they may belong to different categories. Further research is needed to discover whether labels can group only some items which already are quite similar and if labels can segregate relatively dissimilar items. In other words, we still have to understand if there is something like a threshold of similarity which limits the power of labels. It may be that the similarity induced by labels interacts with visual similarity. The discussion of these questions will be in Chapter 3.

Chapter 2

Top-down theories

In the past decades, there have been many attempts to summarise the results and the available theoretical explanations of the effects of labels on perception (e.g., Ferguson & Waxman, 2016; Plunkett, 2010; Robinson et al., 2012; Waxman & Gelman, 2009). The main dichotomy in this debate is whether the effect of labels is top-down or bottom-up. (Plunkett, 2010) uses the terms *supervisory* and *non-supervisory*; Waxman & Gelman (2009) describe two metaphors “child-as-data-analyst” and “child-as-theorist”. These distinctions are not identical, and they stem from different backgrounds. Nonetheless, they all address a similar question: whether the effect of labels stems from higher levels of cognition, or it is merely perceptual.

It seems “natural” to attribute to labels a particular effect because they belong to language and language is known to affect cognition. However, it could also be the case that labels, in virtue of their being also auditory stimuli, contribute to the process of perceptual categorisation. In other words, the effect of labels may be a case of cognitive penetration, but it is also possible that labels and visual stimuli are computed as a perceptual compound.

Despite the different shades this distinction could have, it is useful to distinguish two principal groups of theories; for simplicity, I will call them”

top-down” and “bottom-up”. In this chapter, I will describe and discuss three positions inside the top-down view; in Chapter 3 I will present two positions on the bottom-up side. On the top-down side, the first theory I will discuss is “Natural Pedagogy” (Csibra & Gergely, 2009), the second option is that labels have their effect because labels “refer” (Ferguson & Waxman, 2016; Waxman & Gelman, 2009), the third one is the so-called “Label Feedback Hypothesis” proposed by Lupyan (2012b). On the bottom-up side, two psychologists claim that labels may act as features Vladimir Sloutsky (e.g., Sloutsky & Fisher, 2012) and Kim Plunkett (e.g., GIoZZi et al., 2009; Plunkett et al., 2008).

2.1 Natural Pedagogy

Natural Pedagogy is a theory proposed by the psychologists Csibra and Gergely, according to which human communication is evolutionally designed to transmit knowledge (Csibra & Gergely, 2009, 2011; Gergely & Csibra, 2013). Natural Pedagogy is conceived to be an answer to a particular instance of the so-called *Problem of induction*. The Problem of induction originates from a long-standing philosophical tradition. The core problem is how inductive reasoning can lead to knowledge.

In a psychological perspective, this problem can be declined as it follows: how do we acquire knowledge from a singular instance of information? The usual answer to this question relies on statistics: multiple episodes are sampled to form the basis of generalisation to novel episodes). According to Csibra and Gergely, there is a cognitive short-cut which bypasses repeated exposure to experience: knowledge transmitted via human communication naturally refers to kinds, not only to exemplars. Learners are guided by ostensive signals which make manifest what is relevant and what information could be generalised. In this way, children do not have to be repeatedly exposed to the same stimulus.

Their claim is based on three empirical observations:

1. infants are sensitive to ostensive signals;
2. infants have referential expectations in ostensive contexts;
3. referential communication is interpreted as kind-relevant, and it is generalizable.

2.1.1 Infants are sensitive to ostensive signals

Concerning the first point of their argument, it is proved that infants show receptivity to ostensive signals well before they show evidence of learning.

Ostensive signals enhance infants and children's attention and define transmitted knowledge as generalisable. Even if infants could simply learn by imitation when they are in a communicative environment, learning it is qualitatively different in ostensive contexts. For example, Carpenter et al. (2005) showed that when infants are learning a specific action, if the action is performed with an ostensive signal (e.g., verbally), infants did not only reproduce the final stage of the action but the whole process. When an ostensive signal is present, the transmitted information is taken to be relevant. According to Csibra and Gergely, ostensive signals are a fundamental aspect of human communication. There are different kinds of ostensive signals such as eye gaze, infant-directed speech and infant-induced contingent reactivity.

Eye gaze is the first way children get in touch with ostensive signals. The existing literature about the importance of making eye contact is rather abundant. According to some psychologists, such as Michael Tomasello, the ability to share attention by mutual eye gaze is at the basis of the evolution of cooperation among humans (Moll & Tomasello, 2004). Human eyes are unique: compared to other animals, the size of the sclera is bigger, and it is white. A white sclera allows detecting better the direction of the gaze. Tomasello et al. (2007) proved that infants could understand where someone is watching by following only the direction of the eyes, whereas great apes follow the direction of the whole head. It is only possible to speculate about the developmental trajectory, which led to this effect, but it is a well established – and easy to verify – fact that infants follow eye gaze. The fact that dynamic eye gaze is interpreted as an ostensive signal is supported by recent neuroimaging studies in 4-month-olds (Grossmann et al., 2008).

Another kind of ostensive signal is “infant-directed speech” (IDS), also called “motherese”. It is a well-known fact that newborns prefer IDS over adult-directed speech (Cooper & Aslin, 1990). IDS has a distinctive pat-

tern of intonation, higher and wider pitch, slower speech rate and shorter utterances. Even if it is not strictly ostensive, it is a kind of communication which makes clear that the infant is directly addressed. The same thing happens with gestures: there is a specific modality of making ostensive gestures while addressing children: “motionese”. As for IDS, infants prefer motionese (Kotterba & Iverson, 2009).

Referential expectations

The second part of Csibra and Gergely’s argument is directed to prove that there is a strict link between ostensive signals and referential expectations. According to the authors, preverbal infants are not able yet to fully grasp symbolic modes of reference, such as language or iconic signs. Therefore, at an initial stage, the understanding of ostensive signals is limited to indexical reference, namely to deictic gesture or shifting eye gaze toward them. Older infants have referential expectations for eye gaze. To sustain this claim, they refer to two different studies.

The first one (Csibra & Volein, 2008) shows that infants expect to find objects in the position toward which the gaze is directed. The second one (Moll & Tomasello, 2004) shows that not only infants expect to find objects, but they expect to find objects that belong to the named kind at 12 and 18 months.

Generalisation

The third and last part of the argument is that what infants learn in ostensive and referential communication is generalisable. The latter is the crucial point of this theory: human communication naturally conveys generalisable content.

Ostensive signals indicate that the content has to be generalised and

do not apply only to “here and now”. This is the most delicate part of the entire argument: Csibra and Gergely argue that when communication is ostensive, it is more readily generalised. This is also the most obscure passage in their main article (Csibra & Volein, 2008): their primary reference is a never-published study. This study should show that 18 months old generalise information when it is received ostensively.

Another mentioned study (Yoon et al., 2008) shows that in a communicative context 9-month-olds retain qualitatively different information about novel objects, namely they do not pay attention to the location, but they focus on the identity of the object. To sum up, Csibra and Gergely claim that infants are particularly sensitive to ostensive signals, in ostensive contexts, they have referential expectations, and that when communication is referential, it can be generalised. According to the two psychologists, this mechanism improves learning about generics by making unnecessary repeated exposure to the stimuli.

Critical assessment

There are two different levels at which one may question this theory: its general validity, as done by Nakao & Andrews (2014) or its pertinence for the studies considered in this thesis. At a general level, Natural Pedagogy, as a theory that belongs to the frame of evolutionary psychology, suffers from the very same weaknesses, which are attributed to any theory in this frame. It has the strength of a well-told story, but it is almost impossible to verify. (e.g., Gannon, 2002). Even if we leave on the side the evolutionary claims, there are other good reasons to be critical about this program.

As mentioned before, Natural Pedagogy is built around three main points:

1. infants are sensitive to ostensive signals;
2. when an ostensive signal is present infants have referential expectations;

3. ostensive-referential communication is generalisable.

As for the first point, there are few doubts about the fact that even young infants are interested in human interaction. Therefore it is not surprising that they are also interested in ostensive signals. So far, nothing is surprising in their claims. Their second point is less clear: they claim that in ostensive contexts, infants have referential expectations; two experiments support this position: the first one shows that infants expect objects in the direction of the experimenters gaze and the second one proves that if the object is named they expect to find an object belonging to the nominated category.

The first experiment, if it were possible to demonstrate that the results can be extended to any ostensive signal, proves that infants expect object in the direction of eye gaze (or pointing). The second one shows that when an uttered name accompanies an ostensive signal, infants expect to find an object that belongs to the named category. The second one proves, at best, that infants at 8 months do understand names. It also shows that when names are combined with an ostensive signal, they expect to find the named object in the indicated place.

Here two different two notions of referent are involved. It is not correct to claim that deictic gesture or deictic eye gaze refers in the very same way as words refer. In the first case, no symbol is involved. A deictic gesture or eye gaze are not symbols which stay for something else; they just *point at*. A word does not point at a “general object”, a word refers to a particular kind of object. The word “cat” refers to cats, and the proper name “Mimi” refers to my own cat. A deictic gesture or eye gaze point to any object in the indicated direction; they do not have content.

There is a conspicuous debate on the meaning of words and what a referent is (Speaks, 2019, for a review), but whatever position we endorse, at the minimum we have to recognise that it is part of the understanding of words

being able to identify their referents. It may be true that when there is a combination of an ostensive signal and a label, infants look for the labelled object (if they understand it). Nevertheless, it is also true that when they hear a familiar name, they look for the object even if a deictic ostensive signal is not present (Bergelson & Swingley, 2012). The fact that infants look for the named objects is part of their lexical competence (Marconi, 1997): they can identify referents; this is not special to ostensive contexts. There is evidence that at 6 months infants can already recognise some words, well before speaking (Bergelson & Swingley, 2012). If infants understand words, it means that they can use them *referentially*. If words are merely associated with single tokens, it is not even possible to test if infants do know the word. For instance, if we suppose that a 6-months-old can associate a specific word, “fep”, only with a particular instance of this word, the “fep” she has at home, the only possible way for the word not to be referential is if it is used in association with *that* “fep”. Any other recognised token of a “fep” would indicate that the name refers to a category and not to a single item.

The third point is the most delicate: they claim that what infants learn in ostensive-referential contexts is considered as generalisable. Evidence is scant. The first study they mention is Egyed (2007); they found that when 18-month-olds see an object and a person expressing an emotion related to that object, they usually interpret the emotion as the person’s particular feelings for that object. If the same happens in an ostensive context, they do not interpret the emotion as the reaction of the single person, but as a generalisable emotion, that concerns the value of the object. The fact that an emotion can be perceived as a quality of an object should be further investigated. This study seems not to prove that knowledge in ostensive contexts is generalisable; it rather proves that in an ostensive context the emotion is generalisable. To prove that knowledge about the object is generalisable, it

should be discovered that infants can generalise the emotion to other members of the same category. Unfortunately, this study is not published, and it is impossible to evaluate it.

The second study, Yoon et al. (2008), shows that infants neglect information about the location of an object in ostensive contexts even if they usually pay attention to its location. At best, this study proves that when communication is ostensive, they focus on the object and not on other features.

Both these studies prove that in the presence of a deictic sign infants learn something about the object and not about something else, but none of them proves that what they learn can be generalised to other members of their category.

2.1.2 Natural Pedagogy and the effects of labels in categorisation

In Chapter 1, we examined the existing literature on the role of labels in categorisation in young infants. Some of the results of these studies could be explained by recurring to Natural Pedagogy. In most of the experiments, an experimenter was actually present during the procedure, and she uttered the labels by directly addressing the infant. It could be the case that infants paid more attention because they were inside a social context. (Ferguson & Waxman, 2016) has proposed this theory as an explanation of what happens in the first months of life. This theory could actually explain some of the effects reported in 1.1.6: when infants are 3/4-months-old, they may be affected by ostensive signals and eye gaze.

In 1.2.3, we saw that even if there are many studies on this topic, only two of them actually describe the role of labels in categorisation in infants adequately: Althaus & Westermann (2016) and Plunkett et al. (2008). As reported in Hu's doctoral thesis, (Hu, 2008), in the experiments in Plunkett

et al. (2008), the auditory stimuli were played by a recorded voice. The experimenter was not visible by the infant, therefore no eye gaze was involved. Furthermore, in most of the experiments, the label is embedded in a carrier sentence, such as “Oh look, an X!”, but in Plunkett’s experiments this is not the case. The labels were presented isolated, without a carrier sentence, as described in 1.1.5. For these two reasons, it is not possible to ascribe to Natural Pedagogy the reported effects in categorisation.

2.1.3 Conclusion

In this section, we have seen that even if Natural Pedagogy is an appealing theory, for the purpose of this thesis, it can be criticised at three different levels.

1. it is a theory that belongs to the field of evolutionary psychology, and this field suffers from methodological weakness;
2. the argument that is meant to prove Natural Pedagogy, as reported in 2.1.1, is not sufficiently supported by empirical evidence;
3. it is not an adequate explanation of the only two studies which show the effects of labels.

2.2 The Referential Role of Labels

When looking at the existing literature on the effects of labels on young infants, it is evident that a group lead by the American psychologists Sandra Waxman conducted most of the research on this topic since the Nineties. Their position in the debate remained stable over the years: they claim that labels highlight commonalities and promote categorisation (e.g., Ferguson & Waxman, 2016; Ferry et al., 2010; Fulkerson & Haaf, 2006; Waxman & Braun, 2005; Waxman & Gelman, 2009).

This theoretical statement is supported by over two decades of empirical research with young infants. Most of the experiments share the same setting: a familiarisation phase followed by a novelty preference task., Fig.2.1. The same set of stimuli is usually familiarised in silence, with a word or with a sound. According to their data, the typical result is that there is a novelty preference only when the stimuli are presented with a word, in contrast, there is a lack of preference when the stimuli are paired with a sound.

As we have already discussed, section 1.1, this setting is not adequate to compare the output of categorisation in silence and with a word. The novelty preference task only followed the conditions with an auditory stimulus and not the one in silence. If it is not possible to determine the boundaries of the categories in silence, it is not possible to establish how words impacted it. At best, they managed to prove an advantage of words over sounds: in the same conditions sounds hindered categorisation, which was achieved with a word.











Trial	Word	Tone	Left Screen	Right Screen
Familiarization 1	Look at the toma! Do you see the toma?	---- -- ---- -- ---- -- ----		
Familiarization 2	Look at the toma! Do you see the toma?	---- -- ---- -- ---- -- ----		
Familiarization 3	Look at the toma! Do you see the toma?	---- -- ---- -- ---- -- ----		
Familiarization 4	Look at the toma! Do you see the toma?	---- -- ---- -- ---- -- ----		
Familiarization 5	Look at the toma! Do you see the toma?	---- -- ---- -- ---- -- ----		
Familiarization 6	Look at the toma! Do you see the toma?	---- -- ---- -- ---- -- ----		
Familiarization 7	Look at the toma! Do you see the toma?	---- -- ---- -- ---- -- ----		
Familiarization 8	Look at the toma! Do you see the toma?	---- -- ---- -- ---- -- ----		
Test				

Figure 2.1. An example of the stimuli typically used Waxman and colleagues, Fulkerson & Waxman (2007).

However, we can set aside these methodological issues and focus on their theoretical claims, even if they did not manage to bring empirical evidence for their theories. Also, among the experiments which compare categorisation in silence and categorisation with a word which can be considered methodologically reliable, there is a difference between these two conditions. Althaus & Westermann (2016) and Plunkett et al. (2008) show that labels do shape categories; it could be the case that Waxman and colleagues might have measured a similar effect if they compared categorisation with words and categorisation in silence. Furthermore, their theory is a good candidate to explain other results, such as those of Experiment 5 in Plunkett's study (2008).

2.2.1 The argument in Waxman & Gelman (2009)

Waxman and colleagues support a strong top-down position: labels affect categorisation because they are referents. This claim is clarified in Waxman & Gelman (2009), where an argument to support this claim is built around four main points:

1. Words do not merely associate; they refer. Words are quintessentially symbolic elements.
2. Words and concepts are more than a collection of sensory and/or perceptual features. Even as infants and young children build their lexical and conceptual repertoires, they are also guided by abstract conceptual knowledge (e.g., animacy, intention and cause).
3. Words and concepts are not unitary constructs. There are different kinds of words and different kinds of concepts, and sensitivity to this variety emerges within the first years of life.
4. Words are located within intricate linguistic and social systems. Thus, a word takes its meaning not merely from its history of co-occurrence with entities in the world but also and importantly from the linguistic and social systems in which it is embedded. (Waxman & Gelman, 2009, pp. 258-259).

Their main goal is to oppose associationism as a theory of learning. Associationism has a long-standing tradition both as a philosophical and psychological theory (for a review, Mandelbaum, 2017). The core idea of associationism is that pairs of thoughts become associated on the basis of past experience. Therefore, on a strict associationist account, what counts for words and concept learning is the repeated exposure to the association of a specific label and a specific visual stimulus; this is considered enough to establish a connection between sounds and images and learn a correlation.

What matters for learning is that infants are sensitive to statistical covariation (e.g., Plunkett, 1997; Smith et al., 1996).

According to Waxman and Gelman though, associationism alone is not enough to explain word learning; the metaphor of the “child-as-data-analyst” and the metaphor of the “child-as-theorist” should be equally involved in explaining early word learning and conceptual development. They are convinced that the models proposed by Sloutsky and colleagues (Sloutsky, 2003; Sloutsky et al., 2007, 2001; Sloutsky & Robinson, 2008) do not exclude the possibility that a top-down mechanism plays a role in word learning. Even if Waxman and Gelman declare themselves open to commit to such a double mechanism, their position is that labels can not be features, because labels refer. In the next paragraphs, we will analyse each of the four points of their argument.

Words refer

The first part of the argument is dedicated to proving that words do not merely associate, words *refer*. Because of this property, words are not additional features of the objects. The philosophical debate around reference is complex and articulated (for a review Michaelson & Reimer, 2019). A minimalistic definition of reference should account for the fact that common nouns refer to objects in the world (and not to mental entities, Putnam (1973)) and that they do not refer to specific objects, or *tokens*, as proper names do, and that they refer to categories or *types*.

To support the claim that infants use nouns as referents, they cite a study by Allen & Carey (2004). Allen and Carey trained 18- and 24-month-olds to learn the word “whisk” associated to the picture of a whisk; infants were then asked to extend the name “whisk” either to another picture of a whisk or to

a *real* whisk. They found that infants are more likely to extend the name to the tri-dimensional object rather to the other picture of a whisk. This result is interpreted as proof that words refer to concepts and not to visual stimuli, such as pictures. According to a strict associationist account, infants should have preferred the picture of the whisk because it is more perceptually similar to the first familiarised image.

It is actually possible to interpret this result from a different perspective. By the age of 18 months, infants have started to use their first words, and thus they have been significantly exposed to the fact that nouns *primarily* refer to *real* objects and not to pictures or to other kinds of representation (e.g., plastic toys). Thus, their preference for the real object may not depend on a high-level assumption on the relationship between names and concepts. It could be that they have observed that even when they learn a new word via a picture, its primary referent is the *real* object. It is also possible that at 18 months a picture is already considered as a referential symbol as well, namely that infants know that pictures have real-world referents.

There is no doubt about the fact that words, generally, refer. What should be proven, instead, is that young infants use words as referents in the very same way as adults and the experiments they mention do not go in this direction. Furthermore, it should be explained why having a referential role prevents word form also being associated with visual stimuli.

Early words have perceptual content

The second point of their argument is focused on showing that word learning can not be considered only as mapping linguistic sounds onto perceptual units. They enumerate a series of reasons to prove that words refer to concepts and not to visual stimuli:

1. words are mapped into concepts which have more properties than the visible ones;
2. words often refer to absent things;
3. there are many words which can not be mapped into concrete referents;
4. there are words that do have a concrete referent, but their meaning can not be grasped by observation alone¹.

Everything they list is undoubtedly true, even if it is not clear how this should be relevant for pre-verbal infants ². Nonetheless, it would be possible to postulate a double mechanism: one for words with a concrete referent and one for other words. Also, it is possible that some words with concrete referents are learnt in ostensive contexts, and some other words are learnt in virtue of infants' inferential abilities (see Marconi, 1997). Nothing prevents us from postulating such a distinction. If no word were learnt in an ostensive context, there would be no guarantee of their meaning.

The list mentioned above would be an actual argument against associationism only if someone claimed the opposite. No one claimed so; this is a classic example of a straw man argument: it is an informal fallacy which consists in rejecting an opponent's argument, but the opponent has never proposed that argument. No one ever claimed that every single word is associated with a percept; it would be impossible. Even if words have a perceptual content, again, it does not prevent words from enhancing visual similarity in a bottom-up fashion.

¹To clarify this point, they cite an example from L. Gleitman & Papafragou (2005): if a dog is running behind a cat the very same scene represents, at least, two concepts: CHASE and FLEE.

²From a philosophical perspective it is not correct to say that words refer to concepts, it is a more complex debate. Although there are good reasons to think that the meaning "is not in the mind" (Putnam, 1973), for the present discussion, it is enough to acknowledge that in the meaning of a word there is more than its visual referent.

Words and concepts are not unitary constructs

The very first paragraph of this section of Waxman & Gelman's paper is inaccurate. They report that according to Sloutsky & Robinson (2008), any word can work as an attentional spotlight and facilitate categorisation, but this is not exactly what Sloutsky and colleagues claimed. As described in section 1.1, it is clear that language can enhance attention because of its social role, but only names affect categorisation. No one ever claimed that all words function alike; this is another case of a straw man argument. It is a well-established fact that infants can distinguish different parts of speech; it would be somewhat pretentious to ignore it. Nonetheless, Waxman and Gelman report that Sloutsky and colleagues claim that every word can tune attention and act as a feature.

The social role of words

The last part of the argument stresses the importance of the social context for word learning. According to Waxman and Gelman, the meaning of a word can not depend only on its association to a percept; the relation with other linguistic elements is crucial. This may be true, but it is very unclear how this should be an argument against associationism.

Sloutsky, Plunkett and their research groups did claim that labels can contribute to the overall similarity of compared entities, but they never intended to extend this mechanism further than a laboratory condition. If labels act as features, the way these features interact with an ecologically plausible mechanism of learning is unclear. Infants do not learn words thanks to multiple expositions to the pairing of a name and visual stimuli in a rapid sequence. Infants do not see ten cats while an adult calls "cat" each of them. Nevertheless, there is no reason to believe that even if words are usually learnt in a social environment, it is possible to use them as features in some particular

conditions.

2.2.2 Critical assessment

The idea that labels facilitate categorisation because they refer stems from a common-sense assumption on the role of language. The way Waxman and Gelman describe this theory, though, does not provide a detailed explanation of how this happens.

Their position is rather plausible, despite the lack of the description of a possible mechanism, if we consider children and adults. It becomes less obvious when it is referred to infants.

In section 1.2, we have seen that only two experiments on the role of labels in categorisation in young infants can be considered reliable: Plunkett et al. (2008) and Althaus & Westermann (2016). Therefore, if we want to examine whether the idea that words are special because they refer, we should relate it primarily to the results of these two reliable studies.

The participants of the two studies were 10-month-olds, it means that, even if they could understand some words, they were still pre-verbal infants. In Plunkett et al. (2008), the auditory stimuli were played by a recorded voice, and they were presented without a carrier sentence; it was not a conventional social context. The duration of the presentation of the labels and the visual stimuli was too short to allow infants to learn the new names (Hu, 2008). There is evidence that children by the age of 6 months know the meaning of some nouns (e.g., Bergelson & Swingley, 2012), but in the experimental conditions of Plunkett et al. (2008) infants did not learn the names of the displayed items. In the same way, from a comparative analysis, reported in section 1.1, we know that at 10 months only names, and maybe adjectives, influence categorisation. But the experimental setting, overall, is far from what happens in a regular learning context.

As explained in the Introduction, there is no reason to ascribe to infants in the mentioned experiments more than the ability to perceptually categorise, which does not require referential abilities. To sum up, most of the critiques that Waxman and Gelman propose against associationism are not strong enough to exclude it.

Comparing language and colourful stickers

There is a study where Waxman and colleagues tried to bring evidence against the idea that labels can be associated with visual images and thus facilitate categorisation: Graham et al. (2012). It is an experimental study whose purpose is to show that names are not like other perceptual features. The age group they tested is 4 to 5 years old; this choice makes the result irrelevant for this thesis, but their purpose was to rebut Sloutsky, who often studies similar age groups. Despite these relevant methodological difference, it is crucial to notice that Graham and colleagues decided to compare three category markers: novel nouns, novel adjectives, and colourful stickers. The experimental set-up is rather similar to the ones already mentioned: children were familiarised with a set of stimuli in the three described conditions; in the test phase it was assessed whether one of the category markers had a positive effect on the output of categorisation.

Given the age of participants, it is not surprising that only new names had a facilitatory effect: 4-year-olds are old enough to distinguish the different roles of speech-parts. A core assumption of those who claim that labels also acts as features is that they have to be names; they do have to be linguistic stimuli in order to facilitate categorisation. The bold claim they make is that names can be features even if they belong to human language.

2.2.3 Labels highlight commonalities

One core assumption of the supporters of the top-down positions is that labels facilitate categorisation because they *highlight commonalities*. This is not a verified statement, but rather an unjustified inference. If labels facilitate categorisation and categorisation is based on visual similarity, labels must, somehow, increase visual similarity. The lack of novelty preference in the conditions with sounds is interpreted as the failure of recognising the similarities among the familiarised objects. Surprisingly, this idea has not been further investigated by Waxman and colleagues.

Nevertheless, there are two studies which address this question: Althaus & Mareschal (2014) and Althaus & Plunkett (2016).

Althaus & Mareschal (2014)

Althaus & Mareschal (2014) tested a group of 8-month-olds and a group of 12-month-olds in the same four conditions: Silence, Label (label embedded in a carrier sentence, e.g., “Look at the Timbo!”), No Label (only a sentence, e.g. “Look at this!”), and Sound. Their visual stimuli were designed in a way such that it was easy to keep spatially separate the visual features, Fig.2.2.



Figure 2.2. The stimuli in the upper part are an example of those used in the familiarisation phase; those in the lower part are those used in the test phase (Althaus & Mareschal, 2014).

They found that in 12-month-olds, both labelling and non-labelling phrases facilitate categorisation. The categorisation was not achieved in silence and when images were paired with a sound. 8-month-olds did not categorise the items in any condition.

It may seem that labelling and non-labelling sentences had the same effect in categorisation, but it is only apparent: there was a relevant difference in the patterns of eye gaze during familiarisation. In the Label condition, there was a preference for the shared features of the objects since the beginning of the familiarisation phase. Infants in the No Label condition, instead, showed a preference for the shared features only in the last part of categorisation.

The results of these experiments show that labels increase the attention to the common parts at an earlier stage if compared to the condition without labels. The main issue of this research design is that there is not a direct connection between this attentional tuning effect and increased performance in categorisation.

Althaus & Plunkett (2016)

The experimental of Althaus & Plunkett (2016) is similar to the one of Althaus & Mareschal (2014), but their results go even further in describing how labels direct attention. They had two goals: testing whether individual infant’s categorisation performances are related to the degree to which they focus on commonalities in the presence of a label; control if the commonality focus persists after learning, even when labels are absent. Their stimuli are quite similar to those of the previous study, even if their ecological plausibility is really low: rather than being a single object, they look like two separate objects linked by a string, Fig.2.3.



Figure 2.3. An example of the stimuli used by Althaus & Plunkett (2016).

However, the clear separation of the features made it possible to track infant’s attention during learning. The “leaf-part” had low variability, and the “shell-part” was the most variable part. The interesting part of their results is that labels did not significantly improve categorisation, but labels modified the way categorisation was achieved:

1. when objects are familiarised with a label, infants maintain attention for a longer time;
2. if infants focused more on the common parts of the objects when they were in the Label condition, they were also more likely to show a novelty preference at test; those who focused on the common parts as well, but without a label were less likely to show a novelty preference;

3. at test, infants in the Label condition, paid more attention to the common part of the out-of-category object if compared to infants familiarised in silence.

Labels, here, had the function of identifying the so-called *diagnostic* features; that is, features which indicate category membership.

It is essential to notice that infants successfully managed to categorise both with a label and in silence. This result should not be interpreted as in contrast with all the other studies which seem to prove that labels improve categorisation. There are some set of stimuli that can also be categorised in silence; what matters are the cases in which labels allow the formation of categories that otherwise would not have been formed in silence.

2.2.4 Conclusion

The idea that labels facilitate categorisation because they refer and common names *naturally* refer to categories and not to individuals is the traditional view in this debate. The position held by Waxman and colleagues, though, lacks of a detailed explanation about how this happens, even if a possible explanation of how labels highlight commonalities is available. Furthermore, that fact that names refer may not be relevant to explain the results of Plunkett et al. (2008) and Althaus & Westermann (2016).

2.3 The Label Feedback Hypothesis

Among the top-down theories which aim at explaining the effect of labels in categorisation, there is also the Label Feedback Hypothesis. This theory was initially proposed by the psychologist Gary Lupyan to explain the effects of language on adults. Still, it is an option worth considering when evaluating the effects of labels on infants. As Lupyan (2012b) points out, he did not conceive this theory to explain only the effects of labelling, which however play a significant role in the debate, but it stems from a broader debate on Linguistic Relativity.

2.3.1 The paradox of the effects of language

A review of the studies about language and thought shows that most of them have something in common: the effects of language can be easily nullified, for instance, with a verbal interference task.

For example, Winawer et al. (2007) proved that Russian speakers are faster than English speakers in discriminating some specific shades of colour, but this effect can easily be disrupted. Russian has two terms to describe a set of colour which in English are all called “blue”: “goluboy” (light blue) and “siniy” (dark blue). Russian native speakers are faster at discriminating shades of blue if the two compared colours belong one to the “goluboy” category and the other one to the “siniy” category; vice-versa, they perform like English speakers when the two colours are both “goluboy” or “siniy”.

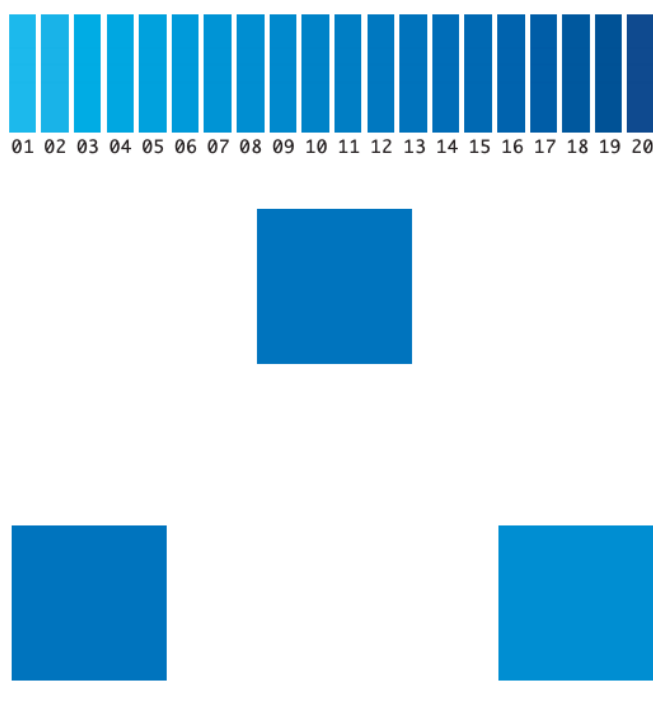


Figure 2.4. An example of the coloured patches used by Winawer et al. (2007).

The advantage of Russian speakers over English speakers disappears when they have to complete the same task while silently rehears digit strings, which is a classic verbal-interference procedure. A similar transient effect can be found in many studies on language and thought (e.g., Athanasopoulos et al., 2015; Cubelli et al., 2011; A. L. Gilbert et al., 2006; Siok et al., 2009; Winawer et al., 2007). A possible way to explain these results is that the perceptual representation of colours is warped by language in long-term experience. Therefore, colours belonging to the same category are treated as more similar. It seems that language can change the perceptual space in “Whorfian” manner.

This property of language, namely modulating perception in an on-line fashion, has received different interpretations. According to some psychologists (Dessalegn & Landau, 2008; L. Gleitman & Papafragou, 2005; Li et al., 2009), the fact that the effects on language can be easily disrupted is

proof that language does not affect thought in a strict sense. This position relies on the assumption that language and concepts are separate entities; in the same way, verbal processing and non-verbal processing are separate. In such a dichotomous perspective, it is hard to explain how language shapes concepts if its effects do not persist. If the two systems are independent, concepts must be permanently modified by language.

Lupyan's theory, the Label Feedback Hypothesis - LFH, has the precise purpose of explaining this paradox. According to Lupyan, language manipulates perception by "manipulating ongoing perceptual processing on-line" (Lupyan, 2012b). Modulation is rapid, automatic, and acts over a distributed interactive system³.

2.3.2 The on-line effects of labelling

In the LFH, the labelling process plays a special role in explaining the effects of language on thought: labels selectively activate the perceptual features that are diagnostic of the labelled category. Lupyan endorses a position according to which categorisation is a process by which "different (i.e., non-identical) stimuli come to be represented as identical in some respect" (Lupyan, 2012b). In this respect, he is an avid supporter of the cognitive penetrability of perception⁴.

When dealing with the effects of labels in categorisation, there are two main theoretical positions: either labels alter perception, or labels alter categorisation, which is not considered as a perceptual process, but a higher-level process. Among those who think that labels alter perception, namely those who think that this is a case of penetration, some claim that early vision is

³See below.

⁴See the Introduction.

penetrated, some others claim that only late vision is penetrated (for a review Raftopoulos, 2019). Lupyan is on the side of those who think that perception is penetrable at any level and he goes even further: he is committed to a collapse between perception and cognition (Lupyan, 2015a,c).

Following Goldstone & Hendrickson (2009), Lupyan claims that the existing empirical evidence is adequate to prove that perception is altered: learning categories warps perception, and it modifies some regions of perceptual space. Categorising objects can not concern only decision making; it is not just about *deciding* what an item is, but it is genuinely a perceptual process. Naming is itself an act of categorisation, and learning how to associate labels and objects is a form of category-training (Goldstone et al., 2001).

According to Lupyan:

“The *label-feedback hypothesis* proposes that language produces transient modulation of ongoing perceptual (and higher-level) processing. In the case of color, this means that after learning that certain colors are called “green,” the perceptual representations activated by a green-colored object become warped by top-down feedback as the verbal label “green” is co-activated. This results in a temporary warping of the perceptual space with greens pushed closer together and/or greens being dragged further from non-greens. Viewing a green object becomes a hybrid visuo-linguistic experience. Knowing that some colors are called green means that our everyday experiences of seeing become affected by the verbal term, which in turn makes the visual representation more categorical. This modulation can be increased – up-regulated – by activating the label to a greater than normal degree as when a participant hears a verbal label prior to seeing a visual display. Conversely, verbal interference is one way to down-regulate the activation of labels leading to reduced influences effect of language on “non-verbal” pro-

cessing.” (Lupyan, 2012b).

This position makes it possible to acknowledge the reversibility of the effects of language and it is committed to a double nature of representation as visual and linguistic at the same time.

The nature of visuo-linguistic representations

According to the LFH, that we should abandon an old model of conceptual representations where there is a distinction between semantic and visual representations. Representations activated by language are multimodal: concepts are not represented by a singular modality, but their representation activate all the modalities involved; for instance, the visual aspects of concepts are represented by some of the same neural structures involved in their visual processing (Barsalou, 2008; Kiefer & Pulvermüller, 2012; Pulvermüller, 2018).

Some evidence about the nature of this hybrid representation is reported in Lupyan & Thompson-Schill (2012). In a series of experiments with a picture verification task, Lupyan and Thompson showed that verbal cues, the word “cat”, had an advantage over non-verbal cues, such as the meowing of a cat, or verbal cue that did not directly refer to the object, the word “meowing”. Conceptual representations are activated by language in a more efficient way. Their findings are incompatible with the idea that labels simply give access to non-verbal concepts because the very same conceptual content should have been accessed in the same way with other cues. Language, presumably, creates different kinds of concepts which may be used for categorising items.

2.3.3 Implications for the studies on infants

Among Lupyan's empirical research, there is a study which is quite similar to those analysed in this thesis, Lupyan et al. (2007), even if it was conducted with adults. They conducted two experiments to investigate the effects of labels on the formations of new categories.

In Experiment 1, participants were told that they were exploring a new planet on which there were two kinds of aliens: those they should approach and those to avoid, Fig. 2.5.

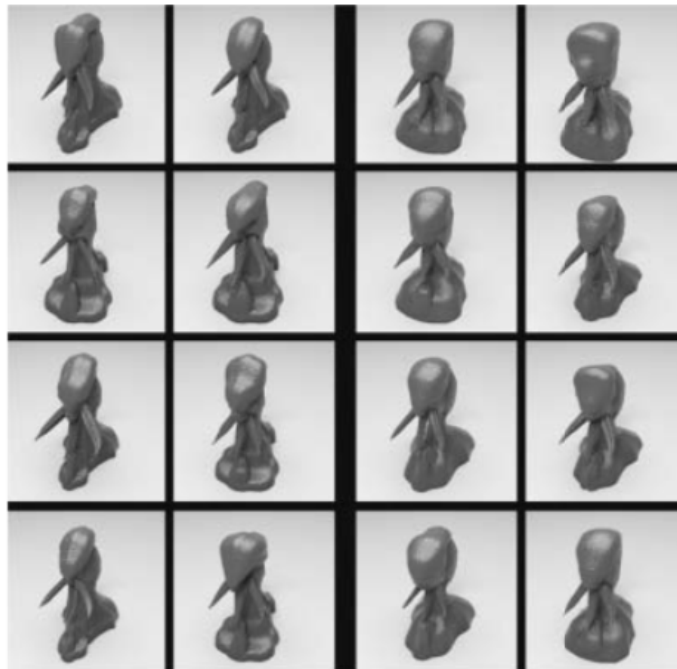


Figure 2.5. The stimuli used by Lupyan et al. (2007).

The aliens were shown one-by-one; after each exposure, participants had to decide to which category the alien belonged and they immediately received a feedback (a buzz for an incorrect response and a bell for the correct one). Participants were split into two conditions, Label and No-Label; for those in the Label condition additionally, after the feedback, a printed label appeared on the side of the alien ("leebish" or "grecious"). After the training trials, in the test phase, an alien appeared on the screen, and the task was to categorise

it as approachable or to be avoided. Among the test stimuli, there were also new aliens not seen in the training phase.

Experiment 2 was almost identical to Experiment one, but the labels were presented in an auditory way. In order to exclude a possible facilitatory role of a feature association which was not linguistic, there was an extra Location condition in which the stimuli were associated with locations and not labels.

As for the results, categorisation was quicker in the presence of labels; the association of a non-verbal feature did not lead to the same result. There is also some evidence that the categories learned with verbal stimuli were more robust as the positive effects of labels lasted even when labels were no longer present and in a follow-up experiment. These experiments corroborate the idea that labels can improve categorisation performances, and, in particular, that labels have a positive effect even when they are redundant, in contrast with the findings of Plunkett et al. (2008).

2.3.4 Conclusion

Connecting these studies, and LFH more broadly, would require some auxiliary hypothesis such as that categorisation in young infants and adults follows the same principles. The present state of research seems to indicate the contrary: categorisation in infants and adults is different, as well as naming categories (Ameel et al., 2008). Whether these differences are negligible requires further investigation. Furthermore, it would be necessary to test whether infants did actually learn labels during the experiments and, if not, it would be necessary to investigate whether there is a feedback mechanism at a neural level even in the absence of a learned label.

Chapter 3

Bottom-up theories

In this chapter, I will present the bottom-up theories proposed to explain the effects of labels on categorisation. Sloutsky & Lo (1999) first proposed the idea that labels are features. Since that early work, the idea that labels may be additional features has received much attention, both on the empirical side and on the computational one.

Traditionally, the labels-as-features account is opposed to the idea, held by Waxman & Gelman (2009), according to which labels affect categorisation because they refer. According to those who support this position, words are supervisory signals that direct and guide learning in a top-down manner. The labels-as-features account, instead, states that labels have a non-supervisory role: “they have the same status as other features and they are handled in the same manner and as part of the same statistical computation as other feature” (Plunkett et al., 2008). Words are part of the stimulus input, and thus they affect cognition in a bottom-up fashion.

The two possibilities do not have to be conceived as mutually exclusive, according to several psychologists (Casasola, 2008; Casasola & Bhagwat, 2007; Plunkett, 2010; Sloutsky, 2010). It is possible that words initially may be non-supervisory signals and later in development, they assume a supervi-

sory role. In theory, it is also possible that labels preserve a non-supervisory role that works alongside with a supervisory mechanism.

If we examine the existing literature, it is rather clear that there is not a shared view of what it means for labels to be features. It is possible to identify two different versions of the theory according to which labels (at least at a certain level) are features, one was proposed by Sloutsky and the other by and Plunkett. Even if they both agree that labels may be additional features, their positions are pretty different, especially for what concerns the development with age of the mechanism and the reasons to support such a position.

3.1 Labels as features in Sloutsky's studies

In a review of the empirical evidence for the role of words in cognitive tasks (Robinson et al., 2012), Sloutsky and colleagues identify two main studies which support the claim that labels act as features: Sloutsky & Lo (1999) and Sloutsky et al. (2001).

Sloutsky & Lo (1999) report the result of three experiments and also propose a mathematical model of similarity: SINC (Similarity, Induction, Naming, Categorization).

3.1.1 The first experiments

The three experiments share the set of the stimuli, which are triads of schematic faces; two of the faces are the test stimuli, and one of them is the target stimulus, Fig 3.1. The faces have three features: the shape of the head, the shape of the ears and the shape of the nose. Each feature has three possible values; for example, curve-lined nose, straight-lined nose or angled nose. The target stimulus can share zero, one or two visual attributes with the two test faces.

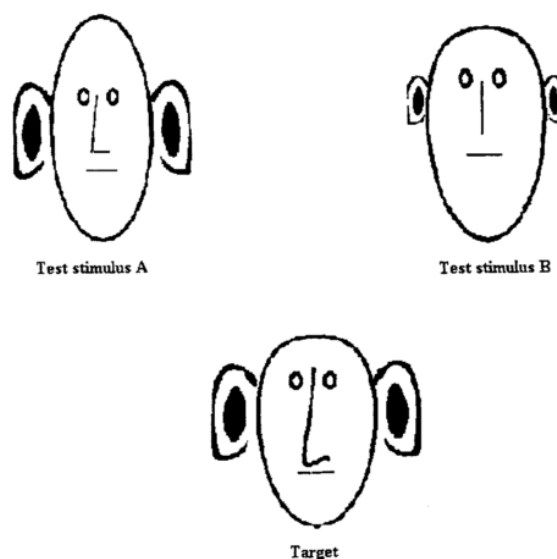


Figure 3.1. (Sloutsky & Lo, 1999, p. 1484)

The stimuli were presented in two conditions: the experimental condition, with a label, and a control condition, without a label. The test stimulus “B”, in the label condition, always shared the label with the Target stimulus, whereas the test stimulus “A” and the target had a greater overall similarity because they had more overlapping features.

In the first experiment, they tested a group of 107 children, aged from 6 to 12, who were divided into three sub-groups based on their age. During the experiments, children were asked which one of the test stimuli (“A” or “B”) was more similar to the target (“T”). The children were tested in a Label condition and a No-Label condition. The stimuli were introduced as pictures of aliens, in the Label condition they were associated with a made-up name. The children were asked to repeat the names, and if they failed in doing it, the experimenter repeated the names once again, to make sure that they learnt it. These were the instructions given:

“I am going to show you some pictures of aliens so you'll learn more about them. Are you ready to start? Let's start! Here we have three alien pictures [pictures were introduced at this point]. They come from different planets (e.g., Guga and Bala). Could you please repeat these names? Look at this one. This is a Guga [points to the target]. This is a Bala [points to Test Stimulus A], and this is a Guga [points to Test Stimulus B]. Is this Guga [points to Test Stimulus B] more similar to this Guga [points to the target], or is this Bala [points to Test Stimulus A] more similar to this Guga [points to the target]?” (Sloutsky & Lo, 1999)

The number of “B” choices was measured in the two conditions (in the label condition, the “B” stimulus shared with the target the same label). The results are displayed in Fig.3.2, where it is possible to see that the discrepancy between the Label and No-Label conditions decreases in the third group with older children.

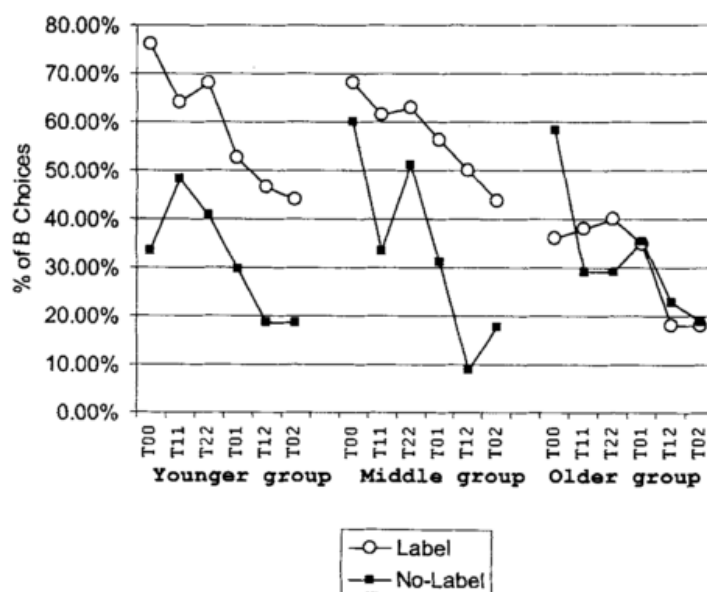


Figure 3.2. (Sloutsky & Lo, 1999, p. 1484)

In the younger group (5 to 7 years old) the “B” choice in the label condition was above chance in the T-0-0 condition (the target shared zero attributes with each of the test stimuli), T-1-1 (the target shared one attribute with each of the test stimuli) and T-2-2 (the target shared two attributes with each of test stimuli). It was at the chance level or below in the other conditions for the younger group and the older group; no significant differences were measured between the younger group and the middle group. It is fundamental to notice that in the younger and the middle group, there is a clear distinction between the results in the label and no-label conditions, but it is not so in the older group.

In the second experiment, there was only one group of children aged 6 to 7 years. The design was the same one of the first experiment, except that the labels were replaced with coloured dots. The dots were introduced as “the colour of alien spaceship”. Thus, instead of telling the children the name of the aliens, there was a red dot beside the face of the alien. The data suggest that although there is an impact of the red dots, the contribution

of labels is significantly greater. The goal of the experiments was to test the *cross-modality hypothesis* according to which the label presented orally is just an additional feature of the visual stimulus. The facilitatory effects on categorisation, according to this hypothesis, depend just on the fact that the input is presented in a cross-modal fashion.

If the effect of labels depends on their being language and not just auditory stimuli, it should be possible to detect it even when language is presented in a non-auditory way. To test this hypothesis, in the third experiment, the labels were replaced by signs of sign language. During the experiment, it was told to the children that the gestural sign was the name of the alien. As for the second experiment, even if it is possible to appreciate the effect of signs, the contribution of labels is significantly greater.

The idea behind this experimental setting is valid: language remains language even when it is not oral. If the effects of language are top-down and they depend on semantics representations, those representations (if they exist) should be activated by language in all its form: oral, written, sign language, written in Braille alphabet, etc. What the experiment neglects is that these different mediums are not equal for everyone. To someone who does not know sign language, it is not language at all, but just signs. In the very same way, for someone who can not read Braille alphabet, it is not language.

It is unclear then why a sign should be for children who do not speak sign language. To be valid, the experiment should be repeated with written language, if children are able to read, or with children who speak sign language.

According to Sloutsky and Lo, the three experiments prove that:

- labels strongly contribute to the similarity judgements of children;

- labels contribute in a *quantitative* manner and these contributions varies depending on the number of overlapping features and age;
- the contribution of labels do not depend on children inability or unwillingness to ignore task-irrelevant information, because they can ignore dots or signs;
- the contribution of labels is larger of the one of non-auditory linguistic entities.

In other words, the reason why Sloutsky and Lo think that labels are additional features is that they did not contribute to the similarity in a *all-or-nothing* manner, but in a quantitative manner. The assumption behind this claim is that if labels had a top-down effect, when a certain item is labelled, it becomes part of a category. If an object is part of a category, it should always be perceived as more similar to the other members of the same category, but this is not what happened in the experiment. Sometimes even when two items are labelled in the same way, the non-labelled items are indicated as the more similar to the Target. If we look at the results of the experiment, it seems that labels contribute to the similarity by interacting with the other features, in a bottom-up way, rather than with a top-down mechanism.

Mathematical model

In the same paper, Sloutsky & Lo (1999) also proposed a formal model, later called SINC (Sloutsky & Fisher, 2004), according to which the similarity of two items (i, j) is given by the equation:

$$Sim(i, j) = S^{N-k} \quad (3.1)$$

where N denotes the total number of relevant attributes (three in the case

of the faces of the previous experiments), k denotes the number of matches, and S ($0 \leq S \leq 1$) denotes the values of a mismatch. If the items are labelled, their similarity can be calculated with the equation:

$$Sim(i, j) = S_{label}^{1-L} S_{vis.attr}^{N-k} \quad (3.2)$$

where S_{label} denotes the values of label mismatch and L denotes a label match. If there is a label match $L = 1$ and therefore $S_{label} = 1$. Following other researchers (Estes, 1994) the value of $S_{vis.attr}$ was set as equal to 0.5. On the basis of existing empirical data, according to which the weight of labels is greater for younger children than for the older, the values of $S_{label/younger}$ was set as equal to $S_{vis.attr}^2 = 0.25$, and the one of $S_{label/older}$ was set as $S_{vis.attr}^{\frac{1}{2}} = 0.71$.

The probability of judging the B item more similar to the test item T, recall that B and T always shared the label whereas A and T did not, is given by:

$$P(B) = \frac{Sim(T, B)}{Sim(T, B) + (Sim(T, A))} \quad (3.3)$$

$$P(A) = 1 - P(B) \quad (3.4)$$

The predictions of these equations have been compared with the obtained results, and they gave a reasonably good fit between predicted and observed probabilities.

3.1.2 Further research: labels and inferences

The first study was followed by several studies conducted by Sloutsky and colleagues on similar topics. After the 1999 experiment, Sloutsky and colleagues moved to a slightly different kind of research questions, namely they

were interested in the role of labels in inductive generalisations. They found in several studies (e.g., Sloutsky & Fisher, 2004; Sloutsky et al., 2001) that labels can drive the inference of new non-visual properties.

If two objects are similar and one of them is known to have a certain property, infants are likely to infer that the other object has the same property too. Sloutsky and colleagues were interested in understanding whether labels could interact or override visual similarity. They found that adults and older children (11-12 years old) tend to consider the shared label as a major indicator of common properties. In contrast, younger children (4-5 years old) evaluate both the physical resemblance and the label.

This field is promising and gives some crucial insight about categorisation broadly conceived. Still, it goes further the purpose of analysing the effects of labels in visual categorisation in young infants. If labels can drive inductive generalisations the mechanism which regulates it would not be a case of cognitive penetration on perception.

3.1.3 Sloutsky's arguments and critiques

To sum up, there are four points which allow Sloutsky and colleagues to state that labels act as perceptual features:

- There is no doubt about the fact that language refers and that a 4 years old child (and probably well before that age) uses words in a referential manner. Even if names refer, they can additionally be considered as features, even for adults.
- The empirical research shows that labels interact with the other features instead of having an all-or-nothing effect. When a child decides what stimulus, "Test A" or "Test B", is more similar to the "Target", she does not rely only upon the overlapping visual features. If the

number of the overlapping features between "Test A" and "Target" is greater than the number of overlapping features between "Test B" and "Target", because "Test B" and "Target" share the same label they are usually considered as more similar.

- Similarly, in the SINC model, the label has a weight that interacts with the weight of the visual features. To mimic the results of the experiments, in the model, the weight of the label has to be kept separate from the other features. It could be interpreted as the fact that labels have a more significant effect on categorisation than single visual features: if, for example, in Sloutsky & Lo (1999), the nose and the label were given the same weight the model would not be able to reproduce the empirical results.
- The label acts in a bottom-up fashion as part of the stimuli, which is a compound of a visual and auditory stimulus) and, thus, it improves similarity as all the other features and not in an all-or-nothing manner.

There are two main critiques which can be moved against Sloutsky's positions; one is about the way they measured similarity, the other one concerns the kind of model they use to explain the phenomena.

In the experiments, to measure perceived similarity, the experimenter directly asked the participant which test item was more similar to the target item. This procedure is not reliable in measuring perceived similarity. Asking someone to judge the similarity of two items does not guarantee that what is measured is what is actually perceived. Furthermore, when testing children and adults, it is possible that they consciously consider the fact that two objects have the same name, and thus when they answer, they are giving to the labels more importance than what they actually have. If the experimenter labelled some items, it is possible that the participants thought that

the label was an even more salient element than what it is. There are some more sophisticated ways to measure the perceived similarity experimentally. For instance, there is the one used by Winawer et al. (2007). The experiment aimed to measure which one of the test stimuli (patches of colour) was perceived as more similar to the target stimulus. Rather than considering the answer itself, that was always correct, Winawer and colleagues measured reaction time. It reflects the assumption according to which if two stimuli are more distant, it is easier to discriminate them, on the contrary, if two stimuli are very similar, it is harder to discriminate¹.

The reason why Sloutsky and colleagues claim that labels interact with the other perceptual features and thus increase perceptual similarity, as previously discussed, is that labels do not act in an all-or-nothing manner. Being called with the same name does not always lead to a positive similarity judgement. Nonetheless, the same data could be interpreted, at least, in another way.

The perceived visual similarity could remain the same and labels lead the decisions about similarity only when the number of common features exceeds a certain threshold. If categorising in children and adults is similar to taking a decision, we could imagine a model where the probability of judging whether an object belongs to a category depends primarily on its visual similarity. If two objects have the same label, but their visual similarity is low the label is ignored, and the probability is low; if two objects have high visual similarity and, additionally, they share the same label, the probability of belonging to the same category is reinforced.

¹For a more detailed discussion about similarity see below, in the General Conclusion.

3.1.4 Conclusion

The way Sloutsky and colleagues consider label as features may not be relevant for the studies with young infants. Still, it is interesting for two reasons: first they offer a good example of an argument for the use of indirect evidence that labels act as features; second, they provide a model to describe the effects which is useful as it gives some insights about how to measure similarity and how labels interact with visual features.

3.2 Labels as features in Plunkett et al. (2008)

The second account which treats labels as features is the one proposed by Plunkett and colleagues. Their reasons to support such a position must be traced back to the experiments conducted in 2008 (Plunkett et al., 2008) and in a subsequent paper in which they offered a theoretical explanation of the previous findings and they replicated the results with a connectionist model (Gliozzi et al., 2009).

As already mentioned, in Plunkett et al. (2008) there were five experiments ²:

- Experiment 1 was conducted with the stimuli belonging to the “Broad Condition”, and Experiment 2 with those of the “Narrow Condition”. The first two experiments replicated the results obtained by (Younger, 1985): after the familiarization phase, the Broad Condition led to the construction of one category whereas the Narrow Condition led to two categories.
- In Experiment 3, the stimuli of the Narrow Condition were paired, consistently, with two new labels (“dax” and “rif”).
- In Experiment 4, the stimuli of the Narrow Condition were randomly paired with the two labels already used in Experiment 3.
- Finally, in Experiment 5, the stimuli of the Narrow Condition were always paired with the same label.

In Experiment 3, the consistent label did not modify the structure of the categories, whereas in Experiment 4 the randomly assigned label disrupted the grouping of the stimuli, leading to a null result in the novelty preference

²For a more detailed explanation of the experiments see Chapter 1

task. In Experiment 5, the use of a single label led to the recognition of the average stimulus as belonging to the category and, therefore, to longer looking time for the extreme stimuli (1111/5555).

Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
Broad Condition	Narrow Condition	Narrow Condition	Narrow Condition	Narrow Condition
Silence	Silence	2 consistent labels	2 random labels	1 label
1 category	2 categories	2 categories	No categorization	1 category

Figure 3.3. Overview of the experiments in (Plunkett et al., 2008)

3.2.1 Interpretation of the results

As discussed in section 1.1 and section 1.2, this experiment is different from most of the experiments on the same topic in a crucial way. It is one of the two cases of reliable procedure which shows how labels impact categorisation. Even if the results of the experiments conducted by Waxman and colleagues have to be dismissed, or only partially accepted, the theory they propose could be valid. If labels act in a top-down manner and facilitate categorisation, what we should expect is that labels improve the performances of infants in Experiments 3 and Experiment 5.

The results of Experiment 5 can fit this hypothesis, but the results of Experiment 3 do not. Experiments 2 and Experiment 3 were both conducted with the stimuli of the Narrow Condition, the only difference is the presence of consistent labels in Experiment 3. If Waxman and colleagues were right, the same stimuli paired with labels should be categorised better. The meaning of “being categorised better” is unclear. However, a possible interpretation is that the average looking time for the new item at test should be longer if compared to the average looking time for the item of the familiarised category.

If we look at the data, that was not the case. It is hard to directly compare the data because in Experiment 2 the test phase lasted 10s, whereas in Experiment 3 it lasted 6s. However, the proportion between the looking time for the object with average values (3333) and the looking time for the objects with extreme values (1111/5555) is almost the same.

Experiment 4 proves that in Experiment 3, infants were able to appreciate the correlation between labels and visual stimuli. In the two experiments, the set of visual stimuli is the same and labels are the same (“dax” and “rif”), the only difference is that the labels in Experiment 4 were randomly assigned, whereas in Experiment 3 they were consistently paired with the two categories of the Narrow Condition.

In Experiment 4, the labels disrupted the formation of the categories: the looking time for the average object (3333) and the extreme objects (1111/5555) was almost the same.

Experiment 5 was conducted in the Narrow Condition, but all the stimuli were paired with the same label. The data are similar to those obtained in Experiment 1, with the Broad Category. The stimuli with extreme values were (1111/5555) preferred to the average one (3333) indicating that the infants formed only one category – as in Experiment 1.

Both Plunkett et al. (2008) and Gliozi et al. (2009) recognise that the results of Experiment 5 are not in contrast with the predictions which stems from Waxman’s positions. It might be the case that labels enhance the similarity between the items and therefore lead to the formation of a unique category.

Labels may have a top-down effect because they refer, things that are given the same name belong to the same category, and things which belong to the same category are more similar. According to the feature-based account proposed by Plunkett, common features do not help in discriminating

between categories, only contrastive features, instead, are relevant for categorical distinctions. For this reason, Plunkett and colleagues claim that in Experiment 3 the labels have an unsupervised feature-based role. In that case, labels were redundant and, therefore, they did not have an impact on categorisation, whereas, in Experiment 4, labels were contrastive and inconsistent.

This is not the only reason to think that labels are features. In a follow-up study of Experiment 3, Hu (2008) provides additional evidence for the non-supervisory role of labels. After the familiarization phase with the Narrow Condition in which the stimuli of the two sub-categories were paired with consistent labels, infants were given an intermodal preferential looking task. In this task, the items 1111 and 5555 (both new, but belonging to the two categories of the Narrow Condition) were shown side by side and each of the two labels was played. If children were exposed to the stimuli (visual and auditory) long enough to learn the association, they should have shown a preference for the displayed objects depending on the label that was played, but they failed to demonstrate any preference. According to Plunkett and colleagues, in Experiment 3, labels had the same role as other visual features and were entered into the statistical computation leading to the category formation.

It could seem that labels play different roles in Experiment 3 and in Experiment 5; GIoZZi et al. (2009) elaborated a unifying explanation of the phenomena. They proposed a model using a neural network, a SOM (self-organising map); the model not only managed to replicate the results of all the experiments, but also predicted infant's behaviour and, in particular, it mimicked primacy and recency effects. Even if the results of Experiment 5 could be explained as a top-down effect, according to GIoZZi and colleagues, they could be explained more easily as a boom-up effect, as described by the

SOM.

3.2.2 Plunkett's argument

To sum up, Plunkett and colleagues have three reasons to believe that labels act as features:

1. In Experiment 3, labels did not impact categorisation. This is what we expect from redundant features, not from top-down effects.
2. The follow-up experiment of Experiment 3 (Hu, 2008) shows that infants did not learn the names of the familiarised objects. According to Gliozzi et al. (2009), a name which was not learnt can not have a supervisory, top-down, effect.
3. The results can be explained with a connectionist model; there is no need to appeal to higher levels of cognition.

I will now evaluate the strength of the first two points; the third point will be discussed in section 3.3.

Contrastive and redundant features

The question of the role of redundant features is reported in Plunkett et al. (2008):

“[...] redundant features do not help discriminate between categories. Contrastive features are the most informative sources in category formation. On this view, labels that do not vary contrastively across sets of objects will be redundant and fail to contribute to category formation”.

The meaning of this passage is not entirely clear, and there are no additional explanations in the article. A possible interpretation is the following.

If we take two identical objects, the number of features that they share is irrelevant: two objects are identical both in the case they have ten features or if they only have two features. A “simple” object is identical to another simple one just as a “complex” object is identical to another complex one. If we consider the case of two objects which are not identical, following Plunkett’s reasoning, contrastive features help to discriminate them. This statement is not obvious.

If two objects have, for instance, a hundred features and the number of overlapping features is ninety-nine, that only mismatching features will not prevent them from being put in the same category. Two cats may differ in the colour of the fur, but this is not a good reason to consider them as belonging to two different categories.

This argument is valid only in a model in which every feature has the same weight, namely a model in which some features are not more salient than others. If we imagine a condition where some features are diagnostic, the mismatch of some other features may not be relevant if there is a match of the diagnostic features. Some features may be more salient than others, such as the shape. Furthermore, there is no general consensus on whether categorisation involves a process of features weighting and features extraction.

Did labels impact categorisation?

Even if we accept that only contrastive features are diagnostic for categorisation, there is an additional problem in this part of the argument. Plunkett and colleagues claim that in Experiment 3 labels did not affect categorisation, and the reason why they claim so is that there were no differences in the duration of the looking time at test in comparison with Experiment 2, as reported in Fig. 3.4.

Table 3
Looking times and statistical analyses for Experiments 1–5

Experiment	1	2	3	4	5
Looking time 3333:1111/5555 (s)	2.76:3.35	3.43:2.72	2.30:1.90	2.14:2.04	1.88:2.14
Percentage of time on average 3333 (<i>SD</i>)	44.17 (9.55)	56.14 (11.81)	55.09 (9.77)	50.28 (12.19)	45.75 (7.72)
<i>t</i> tests (2-tailed)	$t(23) = -2.99$	$t(23) = 2.55$	$t(23) = 2.55$	$t(23) = 0.11$	$t(23) = -2.70$
<i>p</i> value	.007	.018	.018	.912	.013
Effect size (Cohen's <i>d</i>)	0.61	0.52	0.47	0.02	0.55

Note that test trials last 10 s in Experiments 1 and 2 and 6 s in Experiments 3–5.

Figure 3.4. (Plunkett et al., 2008)

In Experiment 2, the percentage of time spent looking the average stimulus (3333) is 56.14; in Experiment 3, the percentage of time spent looking the average stimulus (3333) is 55.09. It is unclear why these data should be an indicator of the lack of effects of labels in categorisation in Experiment 3. Both experiments were conducted with the Narrow Category, which is already composed of two sub-categories. If labels facilitated categorisation, it is unclear what should have happened. One possibility is that infants should have looked at the average item for an even longer time, but, in this case, how much time should have been enough to tell that labels impacted categorisation?

This experimental setting is probably not adequate to measure the impact of labels when they are redundant. There are other ways to measure the facilitatory effects of labels, such as the one used by Lupyan et al. (2007). In that experiment, participants were faster at discriminating two categories if the categories were paired with a label, even when it was redundant.

Non-learnt labels

Hu's doctoral thesis (2008) replicated the same conditions of Experiment 3, and after the familiarization phase, he tested the very same participants to check if they learnt the names presented with the visual stimuli in the

familiarization phase. The infants did not learn the names of the objects. This is taken by Plunkett and colleagues to be a reason to think that the effect can not be top-down, but must be bottom-up.

To be true, this claim requires an additional hypothesis, which is not verified yet, even if it seems plausible: names have to be learnt in order to impact categorisation; besides the fact that maybe names were somehow stored in short-term memory, just like the pictures, but they were forgotten quickly.

If we look at the decrease of attention in the familiarisation phase, as reported in Fig. 3.5, it is easy to notice that in Experiment 1 and 2 infants looked less at the objects, in comparison with Experiment 3, 4, and 5.

Table 2
Mean looking time (s) in the familiarisation trials

	<i>N</i>	Block 1	Block 2	Grand mean/ <i>SD</i>
Experiment 1 (Broad condition)	24	5.810 (1.558)	4.639 (1.297)	5.225 (1.236)
Experiment 2 (Narrow condition)	24	4.337 (1.788)	3.605 (1.672)	3.971 (1.578)
Experiment 3 (Narrow condition)	24	7.298 (1.735)	6.450 (2.052)	6.874 (1.691)
Experiment 4 (Narrow condition)	24	7.438 (1.313)	7.043 (1.365)	7.240 (1.078)
Experiment 5 (Narrow condition)	24	7.386 (1.228)	6.721 (1.602)	7.054 (1.452)

Standard deviation in parentheses.

Figure 3.5. (Plunkett et al., 2008)

They might have enhanced attention, but if directing attention is precisely the way they affect categorisation, it is hard to tell that labels did not impact categorisation even when they were redundant.

3.3 Neural networks as explanations

As seen in the previous section, one of the main reasons in support of the idea that labels are additional features is the possibility of replicating the data with a neural network. The fact that a model can successfully replicate the data, and therefore explain human behaviour, is much discussed. In the next paragraphs, I will briefly discuss the history of neural networks to highlight their role as models for cognitive psychology, and then we will focus on the one that was used by Gliozzi et al. (2009) to mimic Plunkett et al. (2008) results. I will finally discuss whether the model could be considered as an explanation for the empirical data.

3.3.1 Computationalism

Behind the idea that a neural network can describe human behaviour, there is another idea, namely that human cognitive processes are (or can be reduced to) a computation. This idea, widely accepted in cognitive psychology, has a philosophical origin, the so-called “Computational Theory of Mind” (for a review Rescorla, 2017). Describing a process as a computation means thinking that it is an algorithm: “a set of rules that precisely defines a sequence of operations”, as it was famously defined by Stone (1971). Warren McCulloch and Walter Pitts (1943) laid the foundations for modern computationalism. They wrote a famous article, called “A logical calculus of the ideas immanent in nervous activity”, in which they connected Turing’s work on abstract computation, the explanation of cognitive capacities and the mathematical study of neural networks (Piccinini, 2012).

As Piccinini (2012) points out, the importance of McCulloch and Pitts’ account of cognition is that they managed to link neural processes with digital computation, they used mathematically defined neural networks as models, and they supported those models with an appeal to neurophysiological

evidence. After their contribution, three research traditions developed: classicism, connectionism and computational neuroscience.

The dominant paradigm in the 1970s was the classical one. The computer programs developed at the time were designed to simulate human's behaviour despite the psychological plausibility of the mechanisms with which they operate. It is also called "symbolic" because they assumed that cognition is similar to the processing of language-like representations.

Connectionism

Another way to think about the computations underpinning cognitive processes is the one proposed by connectionism. The history of connectionism can be traced back to a seminal paper by Rumelhart and McClelland (1986) about Parallel Distributed Processes. They stressed the importance of parallel processing of neurons and the distributed nature of neural representations. By doing so, they brought the attention again on neural networks after a decade of focusing on symbolic models, linking back to the work of McCulloch and Pitts.

A connectionist neural network is a collection of interconnected nodes; node can belong to three categories: input, output, hidden (which mediates between inputs and outputs)³. If we try to draw an analogy with the brain the input neurons are the sensory ones, the output neurons correspond to the motor ones, and all the other neurons are the hidden ones. Each input unit has an activation value that depends on parameters external to the network. All the hidden units send their activation values to the hidden units to whom they are connected. Then the hidden units calculate their

³The distinction of nodes, also called neurons, in three categories is due to Rumelhart et al. (1986), they introduced the middle, hidden, layer improving in this way the old architectures.

own activation value depending on the values they have received from the inputs. Finally, the signal propagates to the output units to determine their activation values. The activation of the node depends on the weights (also called strength) of the connections between the units. ⁴.

If the information flow in only one direction from inputs to outputs through the hidden layer, it is a so-called feedforward network. It is not a realistic model of how the brain works: the brain has many more connections, and it allows recurrent connections that send the signal from higher levels to lower levels.

3.3.2 Gliozzi et al.'s (2009) self-organising map

The neural architecture used by (Gliozzi et al., 2009) is a self-organising map (SOM). It was used to mimic the results obtained with infants, both during the familiarization phase and at test⁵.

The network consists of a single self-organising map that receives both the acoustic and the visual stimuli as input, as it is schematically represented in Fig.3.6

⁴They can be both positives or negatives; a negative weight indicates an inhibition of the receiving unit. The activation value of each receiving unit depends on an “activating function”, there are many of them, they sum all the incoming signals and if they reach certain threshold they send in their turn a signal.

⁵Before their attempt, other neural networks had already been adjusted to simulate infants' behaviour, and, in particular, categorization and influences of labelling. For a review, see Gliozzi et al. (2009).

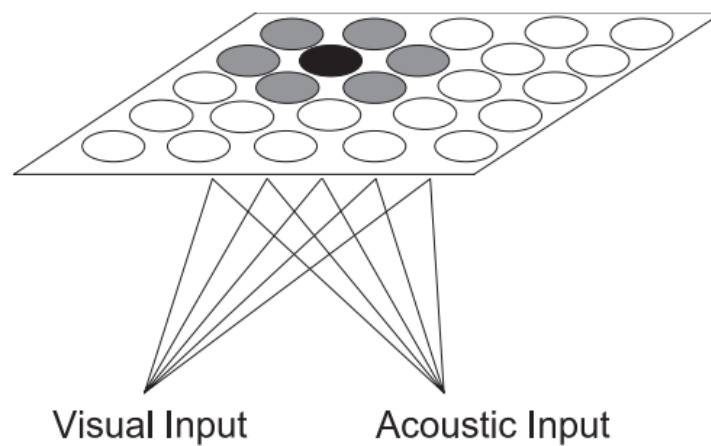


Figure 3.6. The representation of the network provided by Gliozzi et al. (2009).

The network was fed with vectors with four dimensions, both for the visual and the acoustic stimuli. Each value of the input vector carrying the visual information corresponded to a feature of the images used as stimuli by Plunkett et al. (2008): the length of neck and legs, tail dimension and distance between ears. Just like the original experiment, the stimuli can be divided into a Narrow category and a Broad Category. In the Narrow Condition there is a correlation between the dimensions of legs and ears (small values) and neck and tail (high values)⁶. The visual stimuli in the Narrow Condition, from a visual point of view, can be segregated in two categories either for the correlation legs/neck or for the correlation ears/tail, they can be easily visualised in Fig. 3.7

⁶It is worth noticing that even in Younger's original experiments (Younger, 1985; Younger & Cohen, 1986) the numbers of the values used to code the dimensions (e.g., 1111/5555) are not related to the visual dimensions.

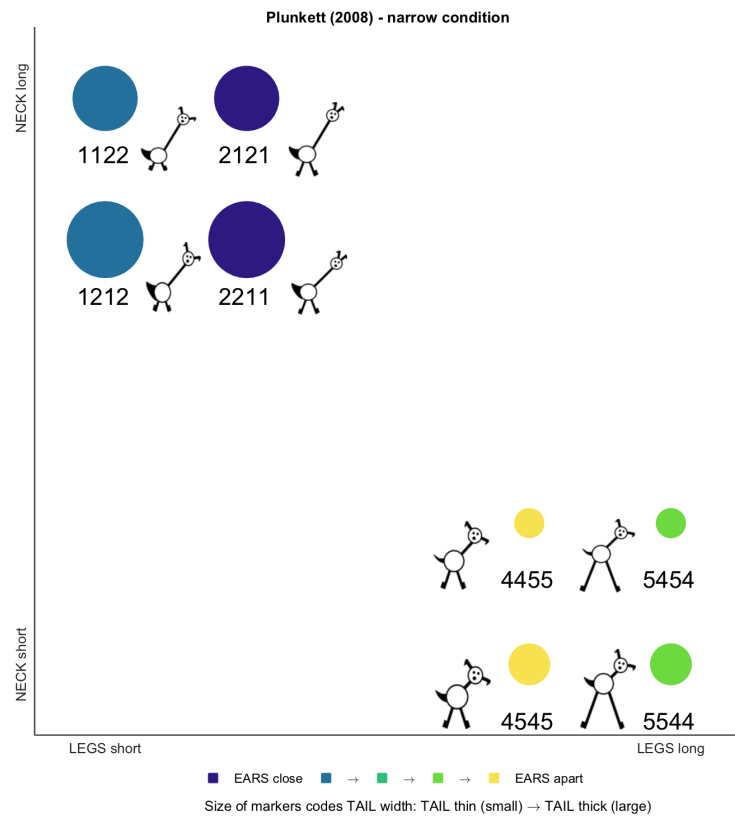


Figure 3.7. Duta 2018.

The first cluster has animals with short legs, long neck, thick tail and close ears. The second cluster is characterized by animals with long legs, short neck, thin tail and distant ears.

On the contrary, the stimuli of the Broad Condition are uniformly spread along the dimensions, Fig. 3.8.

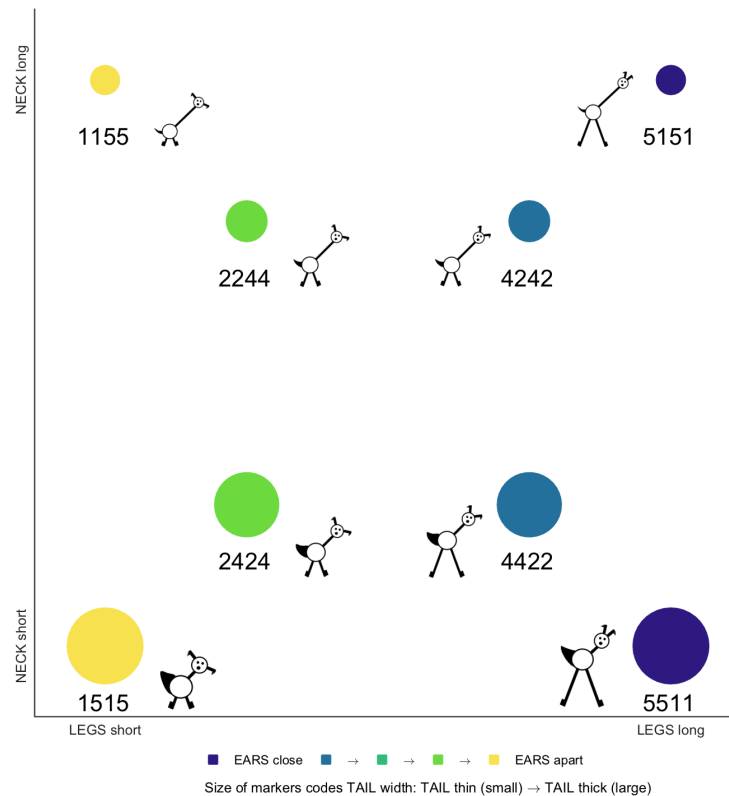


Figure 3.8. Duta 2018.

The visual stimuli are coded following Mareschal & French (2000) who already presented a model with the original stimuli used by Younger & Cohen (1986). Each feature was measured and scaled to range between 0.0 and 1; this reflects the assumption that there are not salient features. For instance, item 3333 was represented by the vector (.64, .62, .61, .56). The acoustic stimuli are coded in a separate vector of dimension four, one label (e.g., “dax”) corresponds to (.0, .0, .7, .7) and the other one (e.g., “rif”) corresponds to (.7, .7, .0, .0). In this model, the visual features are equally salient, and visual and acoustic vectors have the same dimension in order to reflect the assumption that they may be equally important.

Learning algorithm

The SOMs are a special class of artificial neural networks based on *competitive learning*. They use a type of unsupervised learning with which the map learns how to classify the set of inputs by associating similar inputs to similar neurons.

Usually, SOMs are organised in a mono- or bi-dimensional structure. Each element of the input vector is connected to each neuron by a certain weight. The output neurons of the network compete to be activated or fired; in this way, only one neuron at the time is activated. The output neuron that wins the competing process is called *best matching unit* (BMU), or just *winning neuron*. Once it is established which is the winning neuron, it determines the spatial location of a topological neighbourhood of excited neurons. Its weights, and those of close neurons, are modified with the purpose of augmenting the chances of responding again to similar input. Close neurons respond to similar stimuli, in this way, similar inputs are categorised in close regions of space, and therefore the map is organised in a topographic way.

For a standard self-organising map, before learning how to categorise a set of stimuli, the inputs need to be presented more than once to the network. Since this process does not reflect the actual experimental conditions in which infants were tested, the learning rate of traditional SOMs has been adapted to fit this requirement. In the experiment, infants were exposed to each stimulus for 10s only once.

To simulate this condition, two aspects of the learning rate were adapted in the model:

- The learning rate depends on attention, and it is higher when the stimulus is novel⁷.

⁷The novelty of the stimuli depends on the Euclidean distance between them, see Gliozzi

- The learning rate is a function of the total cognitive load: as the cognitive loads increase, the distance between attributes decreases.

These adjustments reflect the idea that if there is less information to process, even small differences can be noticed. Consequently, the values of the learning rate function are higher, for the same Euclidean distance, when the visual inputs are presented alone, without the auditory stimuli. The combination of these two properties allows the model to be fed with each stimulus only once, increasing in this way its psychological plausibility.

Familiarization/training and testing

The same map, consisting of 25 units organised in a hexagonal grid, was used to simulate the five experiments. The net is trained by presenting the inputs in random order; during the test phase, the so-called “modal” stimuli (1111/5555) and average stimuli (3333) are presented to the map, ignoring the acoustic inputs. With infants, categorisation is measured with the novelty preference task; in the model, the membership of a new item to a category is assessed by measuring the quantisation error associated with the test stimuli. According to the authors, the quantisation error indicates the discrepancy between the network’s internal representation and the test stimuli: if it is small it means that the input is not novel, a similar input has already been presented to the model (Gliozzi et al., 2009).

In the experiments, the preference in the test phase was considered as an indirect measure of categorisation; for the network, the quantisation error was considered as an analogue of infant looking times. The quantisation error is a measure of the discrepancy between the network’s representation and the input stimuli. If the quantisation error is small means that for the

et al. (2009).

network the stimulus is familiar because it is similar to a stimulus already presented and vice versa.

Just as the infants, the network exhibits a novelty preference for the modal stimulus if the network looking time (NTL) corresponding to the modal stimulus is higher than the NTL corresponding to the average stimulus. Conversely, if the NTL for the average stimulus is higher than the NTL for the modal stimulus, it is assumed that the network has formed two distinct categories. Recall that in the test phase infants always saw two pictures at the time. The categorisation process was assessed by comparing the looking time for the two displayed items. With the SOM used in the simulation, it was not possible to compare two stimuli simultaneously.

Comparison of the results

The five experiments in Plunkett et al. (2008) were all replicated with a SOM. In Experiments 3-5, both the acoustic and the visual inputs were considered, whereas, in Experiments 1 and 2, the auditory inputs were ignored. For every experiment, 24 networks were trained and tested.

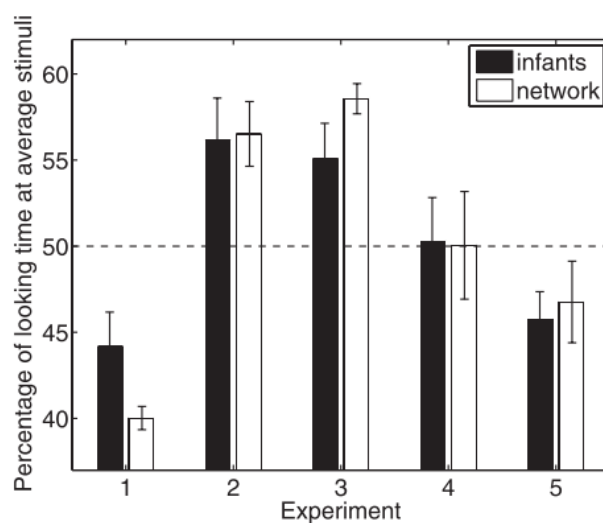


Figure 3.9. The table of results reported in (Gliozzi et al., 2009, p. 724).

Fig. 3.9 shows a comparison of the results of the network and of the experiments. For every one of the five experiments, the looking time of infants was successfully mimicked by the model ⁸. When the label vector was included in the computation, the label had the same impact on category formation that it had on infant's categorization. It is worth noticing that the network, as opposed to infants, in the test phase, was exposed to one stimulus at the time, instead of two simultaneously. Despite this difference, the results obtained with the novelty preference task and by measuring the NTL show a good overall similarity.

When, in the Broad Condition, the network is exposed to the average stimulus (3333), the low NTL depends on the fact that it has close proximity to its BMU in the map. The modal stimuli (1111/5555) do not fit their BMU so closely, and therefore they have a high NLT. The opposite pattern is observed in the Narrow Condition. The presence of the label vector affects the proximity of the testing stimuli to their BMUs. These findings led GIoZZi and colleagues to suggest that "infant preferences are an outcome of on-line processing of the statistical regularities inherent across the set of familiarization stimuli".

Advantages of the bottom-up explanation

Since the networks successfully replicated the empirical results, it was considered as an argument to support the idea that the role of labels on infants could be explained when labels are considered as additional features.

In Plunkett et al. (2008) there were five experiments. The theory according to which labels have a top-down effect can explain the results of Experiment 5, but it can not explain the results of Experiment 3. If labels,

⁸These results are consistent also with Younger (1985).

instead, had a bottom-up effect, this theory would explain the results of all the experiments.

This achievement is made possible thank to the use of the model: it shows that when labels are added as part of the input, they can mimic the results of both the experiments. In other words, there is no need for appealing to a possible top-down role of labels only in Experiment 5, but not in Experiment 3. Everything can be explained with a single theory.

The ability of the model in mimicking the results of the experiments is not enough to consider it as an explanation of those effects. In the next paragraphs, I will consider the psychological and biological plausibility of the model, and I will discuss the predictions it made.

Assessing only the predictions may not be enough: It would be possible that a model makes some right predictions for the wrong reasons. Furthermore, one of the reasons why Gliozzi, Plunkett and colleagues claim that the model explains the empirical results in a bottom-up fashion is that they establish an analogy between how the model is fed with the labels and the supposed role they have in infant's cognition.

3.3.3 Psychological and biological plausibility of the model

In recent years, philosophy started developing theoretical tools for evaluating the goodness of models in cognitive psychology and neuroscience (e.g., Craver & Kaplan, 2018; Kaplan, 2011; Kaplan & Craver, 2011; Ross, 2015). In particular, Kaplan and Craver proposed a criterion to evaluate the explanatory power of models which received much attention called *3M* (model-to-mechanism-mapping).

“(3M) In successful explanatory models in cognitive and systems neuroscience (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism”. (Kaplan & Craver, 2011, p. 611.)

In a strict sense, the model proposed by Gliozzi et al. (2009) does not aim to give a neuroscientific explanation of what happens during categorisation; it is targeted at a high level of cognitive abstraction. Nonetheless, since categorisation is considered as a computational process, and the model is a neurocomputational one, the model should respect similar parameters.

In this perspective, Gliozzi’s model has some features which, in comparison with other similar models (e.g., Mareschal & French, 2000; Westermann & Mareschal, 2004), make it rather plausible.

- It uses an unsupervised learning algorithm instead of backpropagation. The learning algorithms that control the activation threshold of each neuron can be divided into two broad groups: supervised learning and unsupervised learning. A standard version of unsupervised learning is Hebbian learning: when the inputs are presented to the net, if two nodes are active together, the weights between them are increased and vice versa. As already discussed, the learning algorithm used by Gliozzi and colleagues belongs to this group.

Backpropagation is the most widely used supervised learning algorithm. A network that works in this way needs to be trained with a massive set of examples of inputs and the corresponding desired outputs. The network learns because after every repetition the output values are

compared with the desired output, after that all the weights of the net are slightly adjusted in order to match the desired output, it takes a lot of repetitions (hundreds or thousands) before the net learns to produce the correct output. While evaluating the psychological/biological plausibility of a neural network, backpropagation is often a target (e.g., Munakata & O'Reilly, 2000, p. 162.).

- The learning rate depends on attention: it is higher when the stimulus is novel, mimicking infant's behaviour.
- The learning rate is a function of the total cognitive load: as the cognitive loads increase, the distance between attributes decreases. If there is less information to process, even small differences can be noticed, even for infants.
- In its final status, the model has a topological organisation which is similar to the topological organisation of the visual cortex. Similar objects activate close nodes in the map; the brain is known to work according to a similar principle (e.g., Niu et al., 2012).
- The model is fed with each stimulus only once, such as infants in the experiments, but also infants in ecologically plausible conditions. One of the common critiques to neural networks is that they have to be fed with the same stimuli multiple times before learning, infants instead, usually, see the stimuli only once.
- The model is subject to primacy and recency effects⁹.

⁹See below.

3.3.4 Predictions made by the model

Usually, one of the most common parameters to evaluate a model or a scientific theory, in general, is to test whether it can make good predictions.

In the history of philosophy of science, Karl Popper is one of the most famous proponents of predictions (e.g., Popper, 1963). Popper's aim was to distinguish scientific from pseudoscientific theories (such as Marx's theory of history and Freudian psychoanalysis). Pseudoscience was by a vast explanatory power: they could always be adjusted to explain anything. Scientific theories, instead, made predictions about phenomena yet to be observed, and this is what allowed them to be falsified (see Barnes, 2018).

After Popper, Imre Lakatos claimed that *research programs* are empirically progressive as far as their theories predict unexpected facts (Lakatos, 1970). A research program is made of a set of "hard core" propositions and "protective belt" made of auxiliary hypothesis which could be modified to reconcile the data and the hardcore propositions. For Lakatos, pseudoscience coincides with unprogressive programs, namely programs which do not predict new facts.

Nowadays, predictions are still considered as an element to evaluate scientific theories (see Barnes, 2018), in the next paragraphs I will evaluate some predictions made by the model proposed by Gliozzi et al. (2009), and I will also evaluate some accommodations made by the model, in light of the philosophical debate between predictions and accommodations.

Predictions - Experiment 5

The first prediction is that the results obtained in Experiment 5, the one in which the label overrode visual categories, should be only transient. After repeated exposures, the model learns to ignore the label. According to the authors, if the model is correct, infants should show the same effect, namely

after repeated exposures, they should learn to ignore the label and recognise two categories.

In the SOM, learning is not complete after a single exposure to the training data ¹⁰, so, even if initially one single category is formed, the network over time creates a single category. This prediction has never been tested due to the difficulty of engaging 10-months-olds' attention for an adequate amount of time. If this prediction were verified, it would provide further evidence for the role of labels as features. According to the "Labels as Referents" view, the label should instead supervise the categorisation process.

Predictions - Order of presentation and looking time

The model also predicted that categorisation is not only affected by the similarity of the stimuli but also by order of presentation.

The SOMs managed to simulate infant's looking time during the familiarisation phase in two ways:

1. The NLT, just like the infants' looking time, is higher in the first block than in the final one. The quantisation error during the first trials, for the network, is high because the input stimuli do not match the BMUs yet; only after five familiarisation trials, the quantisation error decreases.
2. The NLT, just like the infants' looking time, is higher in Experiment 1 than in Experiment 2, and it is higher in Experiments 3-5 than in Experiment 2.

A possible explanation for this phenomenon is that the stimuli of Experiment 1 have a greater distance from each other if compared to the stimuli

¹⁰Other unsupervised learning architectures do not show this feature. One example is SUSTAIN (Supervised and Unsupervised STratified Adaptive Incremental Network) Love et al. (2004).

of Experiment 2.

For each stimulus presented to the network, GIoZZi and colleagues calculated the distance to the last stimulus previously presented during familiarization by measuring the Euclidean distance between them. The distance is greater for the stimuli in Experiment 1 than for those in Experiment 2 if the new input is compared not only to the previous one but to at least the previous two.

Both infants and the SOMs might be sensitive to the order in which the stimuli are presented. A second analysis of the data in Plunkett et al. (2008) shows that the Euclidean distance, even in the empirical experiment, was higher in Experiment 1 than in Experiment 2. If the Euclidean distance is higher, two stimuli are less similar. The lower the similarity is, the longer is the looking time, both for the network and for infants. As a consequence, the categorization process could be affected by manipulating the order of presentation of the stimuli.

Further work about the order of presentation

The prediction that the order of presentation of the stimuli can affect the outcome of categorisation has been investigated in other three papers: Mather & Plunkett (2011), GIoZZi et al. (2013) and GIoZZi et al. (n.d.).

In Mather & Plunkett (2011), two groups of 10-months-old were tested with a set of stimuli similar to those of (Plunkett et al., 2008), see Figure 3.10. The stimuli were the same, the only difference was the order of the presentation. In the Low Distance condition, the Euclidean distance between each consequent stimulus was minimised and, conversely, in the High Distance condition, it was maximised. So, even if the animals were the same, the similarity between each consequent stimulus was manipulated. The categorisation process was assessed with the novelty preference task.

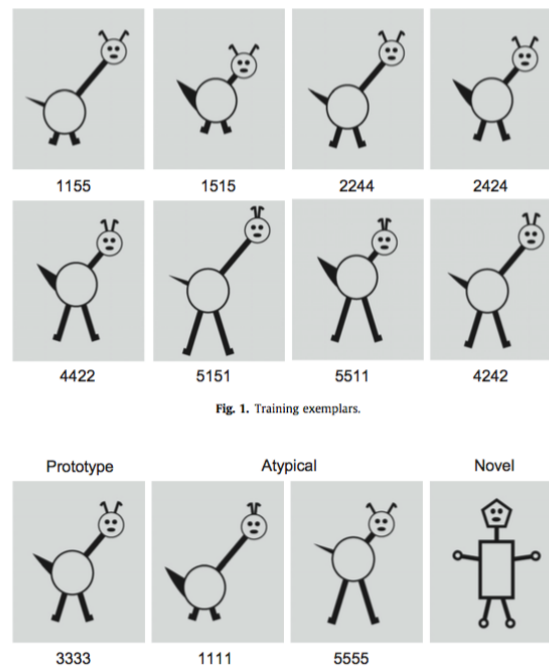


Figure 3.10. Mather & Plunkett (2011)

The test phase consisted of four trials, the first two assessed category formation, the second two assessed infant's novelty preference. In the first two, there was a comparison between the average stimulus (3333) and the two modal ones (1111/5555). In the second two, the choice was between the average stimulus (3333) and a completely novel item, as shown in the lower line of Figure 3.10. If the infants managed to form a single category out of the set of stimuli (independently of whether they were exposed to the Low Distance or the High Distance condition) the average stimulus should be familiar, and therefore the looking time for the modal stimuli should be longer. In case infants do not distinguish between the average and the modal test stimuli, it is necessary to test whether they discriminate between the average or modal and the novel stimulus.

Infants in the High Distance condition exhibited a longer looking time for the modal stimuli compared to the average, those in the Low Distance condition did not show any preference for either of them. There were differences

between the two groups, even in the second two trials: infants in the High Condition showed a stronger preference for the novel item if compared with those in the Low condition. This experiment provides empirical evidence for the idea that categories are influenced by the order of presentation of the stimuli, depending on the manipulation of the Euclidean distance between them.

More recently, Gliozi and colleagues went even further in exploring the effects of order on categorisation. There are several ways to explain the results of Mather & Plunkett (2011), one of them is that infants in the Low Condition might have habituated before forming a complete category, another one is that infants in the High Condition have to “traverse” a larger portion of feature space and it can lead to a stronger representation.

Gliozi et al. (2013) and Gliozi et al. (n.d.) offer a third possible explanation: the first and the last stimuli have a strong effect on the final output of categorisation. This hypothesis has been investigated at first by training SOM’s and at a later stage by testing 10-months-olds. The prediction made by the model performances is close to the performances of infants confirming the existence of primacy and recency effects in the novelty preference task.

This result, taken together with the ones of Mather & Plunkett (2011), prove the validity of Gliozi’s model for what concerns infants categorisation abilities. In this way, it also increments its psychological and biological plausibility. What we can legitimately conclude, as concerns the predictions made by the model, is that it may be a good model *just* for visual categorisation. The claim made by Gliozi et al. (2009) is that labels are merely features and that they interfere in the statistical computation of regularities that is at the basis of categorisation. The prediction about the relevance of the order of presentation (that as we have seen could be a primacy/recency artefact),

does not involve the presence of a label as an additional feature. Thus, it could be the case that the model mimics the effect of labels on categorisation correctly, but this claim does not benefit from the predictions made by the model.

3.3.5 Predictions and accommodations

Traditionally, predictions have been considered to have an advantage over accommodations, the so-called *predictivism*. Predictions are made before the empirical claims are verified and observed. Accommodations, instead, fit an observation which has already been made. A well-supported hypothesis has both predictions and accommodations, but predictions seem to be more impressive. Lipton (2005) gives the example of Halley's comet: Edmond Halley was able to account for the comets of 1531, 1607 and 1682, he claimed that they were a single object with a perturbed elliptical orbit, but only when he predicted the return of the same comet in 1758 its merit was recognised.

The debate over whether predictions should be considered superior to accommodations has been very lively (see Barnes, 2018). However, Lipton (2005) offered a convincing argument in support of the idea that the only difference between predictions and accommodations is a matter of timing.

According to Lipton (2005), there are three main traditional defences of predictivism, I will enumerate them and discuss whether they impact the evaluation of the model used by Gliozzi et al. (2009).

1. *Accommodations allow hypothesis which is built around the data.*

In the case of the model proposed by Gliozzi et al. (2009), this is not an issue. The hypothesis that labels act as additional features was built before the model, and it is supported by independent reasons¹¹.

¹¹Whethere these reasons are strong enough will be debated in the General Conclusion

It can not be considered as an ad hoc hypothesis. Of course, this only partially exonerates the proponents of the model: the idea that labels may be features is independent, but the idea that labels can be added ad that kind of input vectors is not.

2. *Only through predictions a hypothesis gets properly tested.*

This idea relies on the principle that a test is something which could be failed, an idea which suffers from the influence of Popper and the stress he put on falsificationism. Lipton points out that what is true for predictions, namely that if the data had been different, the prediction would have been false, is true also for accommodations. If the data had been different, the hypothesis built around those data would have been false. The accommodation in question here is that labels can contribute to the computation of features and thus affect categorisation. The model presented by Gliozzi et al. (2009) can mimic categorisation in silence, as in Experiment 1 and in Experiment 2. It also shows that if labels are added as additional input vectors, they affect categorisation in the models as auditory labels did for infants: they override visual features in Experiment 5, and they do not improve categorisation in Experiment 3.

Whether labels added as input vectors can replicate all the five experiments in the same way, actually, is one of the weaknesses of the model. The NLT, which measures the discrepancy between the network's representation and the test stimulus, can mimic infant's preferences. Nevertheless, there is another measure which can be taken into account when evaluating the number of categories formed by the model. What the model does is mapping similar inputs into close neurons, the number of categories it forms depend on the clustering technique chosen to measure it: depending on it the model can be considered as having

different numbers of categories. GIoZZi and colleagues used a single-linkage clustering technique, where the Euclidean distances among the input stimuli were considered as the similarity metric, to evaluate the number of categories formed by the model. The number of observed categories depends on the chosen threshold.

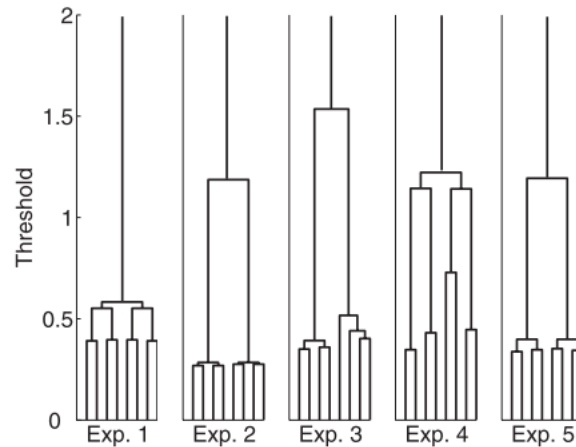


Figure 3.11. GIoZZi et al. (2009).

Fig. 3.11 shows that there is not a threshold which can actually account for the results of all five experiments.

Furthermore, from the very same analysis, the opposite could be derived. If we compare the number of clusters in Experiment 2 and Experiment 3, it is evident that in Experiment 3, were the only difference was the vector carrying the label, two categories can be still present with a higher threshold. This result can also be interpreted as the fact that redundant labels affect categorisation as they make the two categories even more distant.

It is nonetheless true that the model captured the behaviour of infants during the familiarisation phase.

3. *The argument from the best explanation.*

If we consider accommodations, it is possible to indicate two reasons

why the hypothesis fit the data: either the hypothesis is true, or it was designed to fit the data. In the case of predictions, instead, the idea that they were built to fit the data is out of the question. Lipton recognised the considerable intuitive force of this point, but he claims that it also has a number of weaknesses, in particular, the fact that a hypothesis was designed to fit the data does not weaken the inference from the fit to the correctness.

As for the first point, the model is meant to replicate the data of experiments with infants; in those experiments, there were already reasons to believe that labels may be additional features.

3.3.6 Conclusion

A central part of the argument used in Gliozzi et al. (2009) to claim that labels are additional features is that the results in Plunkett et al. (2008) can be replicated with a neural network. In this section, I described the architecture of the SOMs, and I tried to evaluate it on the basis of its biological plausibility and the predictions and accommodations it made. The predictions made by the model about the order of presentation indicates that it is a good model of infant's categorisation. However, the assumption that language can enter the computation as other features did not provide any prediction. If we look at accommodations there is an important remark: even if labels added as input vectors can mimic infants' categorisation behaviour, they do not do it in the very same manner, as explained in the previous paragraph.

Chapter 4

General Conclusion

The purpose of this thesis was to examine the effects of labels on infants and to discuss the existing theories to explain such results. Even if the literature on this topic is abundant, and there are some reviews (Ferguson & Waxman, 2016; Plunkett, 2010; Robinson et al., 2012), most of the studies were not compared yet. In particular, no one already discussed all the possible theoretical explanations.

4.1 Overview of the advancements

4.1.1 The role of labels

In Chapter 1, we have achieved some important results. The first one is that the effect of labels can be ascribed precisely to labels on not to something else, at least at a certain age. This may seem trivial, but even the most recent experiments kept comparing labels and sounds to ensure that labels were responsible for affecting categorisation. Furthermore, it was still unclear whether the effects depended on labels or language in a broad sense. This point is crucial for the whole thesis.

Chapter 2 and Chapter 3 presented two different ways of interpreting the effect of labels on infants. Those who support the idea that labels have a top-down effect also believe that *language is special*, therefore, if it has an effect it can not be a mere association of stimuli which enter the same computation. Those who claim that labels may act as features, at least under certain circumstances, on the other hand never claimed that any association between a visual stimulus and an auditory stimulus could produce the same effect. What is clear from the first chapter, which is also what makes the idea that labels can be features a fascinating challenge, is that labels *are* special and, nevertheless, they could act as perceptual features.

Besides the experiments presented in section 1.1, which were focused on infants, there is other evidence that labels impact categorisation in a special manner. In particular, there are two studies reported in Chapter 2: Graham et al. (2012) and Lupyan et al. (2007). They both compared the effects of labels with the association of other stimuli, such as colourful stickers and non-linguistic sounds. As it was to be expected, only labels had a positive effect on categorisation.

The open challenge for those who claim that labels may be features is to explain how labels can have a bottom-up effect if this is the only association which works: the association with other stimuli, visual or auditory, did not lead to a positive result. It would be necessary to posit a mechanism which selects labels, and in particular names ad salient stimuli for categorisation.

The second advancement is that it is possible to define what this effect is; in particular, it is possible to identify two different effects: a grouping effect and a categorisation effect. Labels seem to increase or decrease perceived similarity, and thus they keep objects together in the same category, or they split them into different categories.

Further directions

One crucial point, which should be further empirically investigated, concerns the amplitude of these effects. In the two studies I focused on in section 1.2, the stimuli had the property of being rather similar. Measuring similarity is an open challenge, as I will discuss further on. Still, even without a precise measure of it, it is not implausible to state that the stimuli used by Plunkett et al. (2008) and by (Althaus & Westermann, 2016) could be considered as belonging to the same basic level; even the stimuli of the Narrow Condition in Plunkett et al. (2008), which in silence are split into two categories.

Labels can slightly increase or decrease similarity of objects which are already quite similar. Would the same happen with objects which have a lower similarity? In other words, it seems that labels can increase the similarity of a set of cats, can a label, such as “mammal”, increase the similarity of cats and cows? A partial answer comes from the experiments of Waxman and colleagues: their sets of stimuli were more heterogeneous. For instance, if we consider the set of dinosaurs of Fig. , we can say that the perceptual distance among them is higher than the distance among the stimuli used by Plunkett et al. (2008) or by (Althaus & Westermann, 2016). Their experiments seem to indicate that labels can also group this kind of stimuli, but since they did not include a Silence condition, there is a lack of reliable evidence.

This is a topic which deserves further investigation. Whether a label could make *any* objects look more similar is an open question.

4.1.2 Top-down theories

Chapter 2 presented the top-down positions inside the debate: Natural Pedagogy, Labels as Referents, and the Label Feed Back Hypothesis.

Natural Pedagogy

As we have seen, Natural Pedagogy is not a good candidate as an explanation of the effects of labels. It is not sufficiently supported by empirical evidence, as a general theory of learning, and, in particular, it is not suited to explain the results of Plunkett et al. (2008) and Althaus & Westermann (2016). Even if it were a valid theory of learning, it would have troubles in explaining the results of experiments where human interaction is minimised.

Labels are Referents

The second account presented in the second chapter is the one of Waxman and colleagues, who are also the researchers who produced most of the existing literature on this topic. According to them, labels facilitate categorisation because labels are nouns and nouns refer; in particular, common nouns refer to categories. Their arguments, though, are weak: they successfully manage to show that names are referents even for infants, but the fact that labels are referents does not imply that they impact categorisation *because* they are referents.

Waxman and colleagues often claim that labels highlight commonalities, even if they do not have direct evidence for it. Two studies, Althaus & Mareschal (2014) and Althaus & Plunkett (2016), explored this possibility, and their results indicate that this may be a growing research field.

The empirical literature on Linguistic Relativity is proliferating; some of the most recent theoretical and empirical studies highlight the role language plays in directing attention.

Slobin (2006, 1987) proposed a theory called “Thinking for Speaking” based on the observation of how children described some images representing a story. The same images for speakers of different languages were interpreted

as different stories. These differences, according to Slobin, depend on the way each language codes the events; it does not mean that language rigidly shapes thought, it just means that when we speak, depending on the language we are using, we pay attention to different aspects of what we see.

On the experimental side, Athanasopoulos et al. (2015), for instance, showed that German-English bilinguals describe motion events according to the grammatical constraints of the language in which they operate. Depending on the language the participants were using, they focused on different parts of the images. It may be the case that even names can direct attention on common features in a categorisation task in the same way language directs attention in other tasks.

Attention is also a target in the debate about the cognitive penetrability of perception (e.g., Gross, 2017; Lupyan, 2017a; Marchi, 2017; Stokes, 2018; Summerfield & Egner, 2009). Whether selective attention represents a genuine case of cognitive penetration is much debated. The case of labels and categorisation could turn out to be an example of such a mechanism.

Label Feedback Hypothesis

The Label Feedback Hypothesis is a good candidate to explain the effects of labels on categorisation with adults. A positive aspect of it is that it offers an explanation of the disruptive effect of verbal interference tasks. Furthermore, the idea that there is no distinction between linguistic and conceptual representations is promising and deserves further investigation.

Whether this theory could also explain the results of infants is an open question.

4.1.3 Bottom-up theories

For what matters the bottom-up theories about labels and categorisation, we have seen that there are two main positions: the one of Sloutsky and colleagues and the one of Plunkett and colleagues. We have also seen that these positions are somewhat different even though they both advocate for the role of additional features for labels.

One main difference is the age of the participants. Sloutsky works with children and Plunkett's experiments were conducted with 10-month olds.

What they have in common is the kind of reasoning that the use to argue that labels act as features. It is virtually impossible to tell based on neural evidence what happens during categorisation, both because imaging techniques are not sophisticated enough and because it is not the kind of test that could be done with infants. The only way to prove that labels act as features is to define what features do and test whether labels work similarly.

Features for Sloutsky

Sloutsky identified the role of features as elements which interacts to contribute to categorisation. The value of a single feature does not determine the belonging to a category or another one; in the same way, when some items were labelled, the label did not automatically place the item in one of the two possible categories of Sloutsky's experiments. Sometimes labels were ignored, and visual features determined categorisation.

For this argument to be true, categorisation should be a model where each feature has a weight, and they all contribute to categorisation. This may be true, but it is certainly not reflected by Sloutky's model of similarity.

$$Sim(i, j) = S_{label}^{1-L} S_{vis.attr.}^{N-k} \quad (4.1)$$

In the equation 4.1, it is evident that, for what concerns visual attributes, the only thing that matters is the number of mismatches given by $N - k$ where N is the total number of features and k is the number of matching features. This is not a model where features interact, a possible model where it happens are the SOMs used by Plunkett.

Furthermore, to fit the empirical data, the similarity of the label (match or mismatch) has to be kept separate, this may indicate that labels have a higher saliency compared to other features.

As already explained in section 3.1.3, these results could also be explained by a probabilistic model where visual similarity indicates the a priori probability of belonging to a specific category, and the match of the label represents additional evidence. Only if the a priori probability is not over a certain threshold the label overrides visual attributes.

Features for Plunkett

Plunkett's argument is based on the results of Experiment 3: redundant labels did not affect categorisation, just like redundant features, and a computational model replicates these results. This argument has already been analysed in details in the previous chapter; to be valid, some points need further evidence:

- The novelty preference task may not be ideal for testing the effect of redundant labels.
- The predictions of the model did not include language.
- The accommodations of the model do not perfectly mimic infants results.

Despite these issues, Plunkett's position is still the best available on the bottom-up side of the debate. Its strength will be discussed in the next section.

4.2 Discussion of the theoretical options

As we have seen, Plunkett's position challenges the traditional idea that labels facilitate categorisation because they refer. To assess whether labels can act as features, the argument must consider the following point:

1. It must be defined what it means to be top-down and bottom-up, to clarify, in both cases what we expect from labels.
2. There must be evidence that labels, in the experiments, independently from the model, act as features.
3. The model must be evaluated, and it should be assessed whether it is enough to exclude that labels facilitate categorisation in a top-down way.

4.2.1 Top-down and bottom-up

The first distinction among the possible theories is between the bottom-up and the top-down ones. In the previous chapters, I broadly explained the difference between these two options, but it is worth noticing that there is no general consensus about this distinction.

Engel et al. (2001) identify four different "flavours of *top-down*", which is an ambiguous concept:

1. *Anatomical*. According to the authors, this is the most widely used variant of this distinction. Top-down refers to the activity of feedback connections in a precessing hierarchy; bottom-up instead refers to the information flow. For example, information going from the retina, through the thalamus, to the primary visual cortex (V1) is bottom-up. Vice-versa, the information going from the secondary visual cortex (V2) to the primary visual cortex (V1) is top-down.

2. *Cognitivist*. In this perspective, the difference between top-down and bottom-up refers to the difference between hypothesis/expectation-drive processes and stimulus-driven processes.
3. *Gestaltist*. This way of conceiving top-down vs bottom-up processing describes the cases where the *whole* determines the perception of the *parts*. The literature on this specific effect is quite broad and is at the basis of gestalt psychology. The main idea is that there is a top-down influence on perception, which induces a “global-precedence”.
4. *Dynamicist*. The last account does not require a fixed anatomical or functional hierarchy. Still, it relies on the existence of flexible recruitment of brain regions for different tasks (which may have a specific function but can also be temporarily converted). It is the case where large scale dynamics can influence local neuronal behaviour.

These four possible ways of conceiving the top-down vs bottom-up distinction do not have to be conceived as mutually exclusive; they just apply to partly different levels of description.

Rauss & Pourtois (2013a) argue that all these possible distinctions are somewhat problematic and that the debate can not be reduced to a simple dichotomy which is misleading both in psychology and in cellular neuroscience. The distinctions outlined by Engel et al. (2001) can be applied to the theories exposed in this thesis.

They wax Waxman and colleagues intend the effect of labels is just that there is an influence of higher levels of cognition on categorisation. Lupyan, who is on the same side of the debate, proposes a top-down effect at a neural level. The distinction made by Plunkett and colleagues is mixed: on the one hand, they propose a distinction between *supervisory* and *non-supervisory* which belongs to the same level of analysis of Waxman and colleagues; on

the other hand, the model proposed shifts the discussion on a properly neural level.

The question is whether it is possible, despite these differences, to define what labels should do in case their effect is top-down and what labels do in case the effect is bottom-up.

If labels act in a top-down manner, we should expect them to direct categorisation: if they act in a bottom-up manner, we should expect them to contribute to categorisation as other features. Therefore, any theory aiming to explain the effect of labels should be able to discriminate between these two options.

4.2.2 Labels and features

The second point of the argument depends on the definition of what it means to act as other features. In section 4.1, we have seen how Sloutsky and Plunkett considered the role of perceptual features: features contribute to similarity; they do not act in an all-or-nothing manner; redundant features do not affect categorisation.

Two separate questions need to be evaluated: (1) whether it is true that this is the role of visual features and (2) whether labels actually assumed the role of features.

As for the role of features, not much can be said: little is known about the way we *perceive* similarity and what role each feature has in contributing to similarity, even if there exist a lot of models (for a review Enflo, 2020).

If we assume that the role of features is the one indicated by Sloutsky and Plunkett, and we consider whether the experiments in Plunkett et al. (2008) satisfy these requirements, there are still some open questions.

Plunkett's argument is entirely built on the results of Experiment 3 in

comparison with Experiment 2: redundant labels did not impact categorisation, therefore they were features. In Chapter 2, I questioned whether the experimental procedure that was used, the novelty preference task, is suited to observe a possible facilitatory effect of labels.

Another argument for labels as features

There is another possible path to show that labels may act as features which was not considered in the mentioned studies.

Features can vary in their dimension: For instance, if we consider the stimuli used by Plunkett et al. (2008), the legs and the neck can assume different lengths, the ears can be more or less distant, the tail can be more or less thick. The same is true for most visual features.

If labels were acting as features, we might argue that we expect labels to vary in the same way. A possible way to describe this idea is that different pronunciations of the same word could affect categorisation in the same way features do.

To test this hypothesis, we could imagine an experiment with a set of stimuli which in silence is considered as one category, such as the stimuli used by Althaus & Westermann (2016). We can then imagine two labelling conditions: one with two Distinct Labels, such as “dax” and “rif”; and one with two Similar Labels, such as “dax” and “tax” or “dax” and “dex”.

If in the Distinct Labels conditions the stimuli are split into two categories and in the Similar Labels condition not, it would provide additional evidence for the theory that labels can be features.

A possibility is that children already ignore small differences in the pronunciation of words because this is how language works (see J. F. Werker & Fennell, 2004). If it were the case, the idea that labels can be additional features would be even weaker.

4.2.3 The model

As explained in section 3.3, when evaluating a computational model there are at least three parameters to consider: the model's plausibility, which is how well the parts of the mechanism of the model can be mapped into the part of the biological and psychological mechanism; the predictions made by the model, and the accommodations it makes.

Plausibility of the model

The SOMs, compared to other kinds of networks, have some characteristics which make them particularly suited to replicate human categorisation, as already discussed. Nonetheless, there are some additional remarks which can be presented. If we consider the stimuli used by Plunkett et al. (2008) in the Narrow Condition, Fig. 4.1, it is hard to claim that all the visual features had the same weight, which is one of the assumptions of the modelGlozzi et al. (2009).

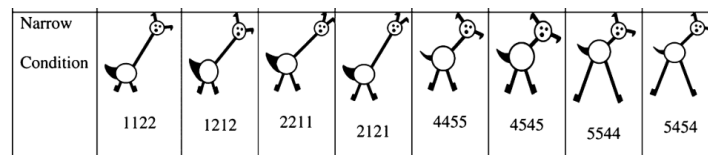


Figure 4.1. The stimuli of the Narrow Condition in Plunkett et al. (2008).

A naive observation which can be done, but should be tested, is that the necks and the legs have a greater saliency than the ears and the tails. At first sight, it is possible to discriminate two categories because the correlation between neck's length and leg's length is very evident; it is harder to notice the dimension of ears and tails. Furthermore, if the dimensions of the circles representing the body of the animals are all equal, the animals with the long neck look like their body is smaller. This gestaltic effect is somehow similar to the famous Müller-Lyer illusion (Bermond & Van Heerden, 1996).

Both these observations should be tested, if they turn out to be valid, the stimuli should be coded accordingly. It may turn out that the two categories are even more distant and that the effect of labels in Experiment 5 is greater than what it seems.

Predictions

Concerning the predictions of the model, the main critique was already exposed in section 3.3: the model made some impressive predictions, but none of them concerned the assumption that labels could be additional vectors.

Accommodations

Even the accommodations of the model have been evaluated in section 3.3, the result of the analysis is that claiming that the model mimics infant's categorisation with labels is a matter of interpretation.

4.2.4 Final remarks

There is one last element which should be considered in this discussion. The distinction between top-down and bottom-up, from a psychological perspective, is less clear than what it seems, as explained above. A possible way to define the role of labels is to try to understand whether they *direct* categorisation or whether they *contribute* to categorisation. In the model proposed by Gliozzi et al. (2009), the role of labels is said to be bottom-up because they enter the computation that leads to the formation of categories and, possibly, because in computational neural networks there is a well-established distinction between top-down and bottom-up processes depending on the network's architecture. The model discussed so far is bottom-up because labels are added as a vector in the same way as visual stimuli. But, we are far from claiming that there is an analogy between what is bottom-up for a network

and what is bottom-up for the brain: there should be a better mapping of the elements of the brain into the elements of the model, which has not been reached yet.

Given these premises, one could question why in the model language, even if considered only as an additional feature, had to be carried in a separate vector which also has a greater dimension compared to the vectors of the visual stimuli. If labels are features, a possibility was to create vectors with an additional dimension which carried both the visual and the auditory part of the input. The fact that the vector is separate seem to reflect the observation that language actually direct categorisation and do not contribute to it.

This point as crucial as it is delicate: to obtain effects similar to infants' behaviour while being part of the computation labels had to be carried on separate vectors and to have a greater dimension. This reflects the idea that even if labels interact with other features, their weight is such that they direct categorisation.

This observation reflects the idea that *language is special* and it is corroborated by the analysis done in the first chapter: only labels affect categorisation, other auditory stimuli do not.

To conclude, the observations made by Plunkett et al. (2008) and GIoZZi et al. (2009) challenge the traditional idea that labels impact categorisation in a top-down manner. In particular, their best objection is that label can achieve this result even when infants did not show signs of having learnt the labels. However, it may be that the experimental setting of Plunkett et al. (2008) was not adequate to show the effect of labels when they were redundant and the model proposed by GIoZZi et al. (2009) does not exclude that the effects of language on categorisation can be top-down.

References

- Allen, M., & Carey, S. (2004). Do both pictures and words function as symbols for 18- and 24-month-old children? *Journal of Cognition and Development, 5*(2), 185–212.
- Althaus, N., & Mareschal, D. (2012). Using Saliency Maps to Separate Competing Processes in Infant Visual Cognition. *Child Development, 83*(4), 1122–1128.
- Althaus, N., & Mareschal, D. (2014). Labels Direct Infants' Attention to Commonalities During Novel Category Learning. *PLOS ONE, 9*(7), 1–10.
- Althaus, N., & Plunkett, K. (2015). Timing matters: The impact of label synchrony on infant categorisation. *Cognition, 139*, 1–9.
- Althaus, N., & Plunkett, K. (2016). Categorization in infancy: labeling induces a persisting focus on commonalities. *Developmental Science, 19*(5), 770–780.
- Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. *Journal of Experimental Child Psychology, 151*, 5–17.
- Ameel, E., Malt, B., & Storms, G. (2008). Object naming and later lexical development: From baby bottle to beer bottle. *Journal of Memory and Language, 58*(2), 262–285.

- Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition*, *128*(2), 214–227.
- Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Scharner, A., Kibbe, A., ... Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, *26*(4), 518–526.
- Au, T. K.-F. (1983). Chinese and english counterfactuals: The sapir-whorf hypothesis revisited. *Cognition*, *15*(1), 155 - 187.
- 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.
- Barnes, E. C. (2018). Prediction versus accommodation. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617-645.
- Beck, J. (2018). Marking the Perception–Cognition Boundary: The Criterion of Stimulus-Dependence. *Australasian Journal of Philosophy*, *96*(2), 319–334.
- Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. In (Vol. 109, pp. 3253–3258). National Academy of Sciences.
- Bermond, B., & Van Heerden, J. (1996). The Müller-Lyer illusion explained and its theoretical importance reconsidered. *Biology and Philosophy*, *11*(3), 321–338.

- Best, C., Robinson, C. W., & Sloutsky, V. M. (2011). The Effect of Labels on Children's Category Learning. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 3332–3336.
- Bloom, A. (1981). *The linguistic shaping of thought: A study in the impact of language on thinking in china and the west*. Hillsdale, N.J.: Lawrence Erlbaum Associate.
- Booth, A. E., & Waxman, S. R. (2006). Déjà vu all over again: Re-revisiting the conceptual status of early word learning: Comment on smith and samuelson (2006). *Developmental Psychology*, 42(6), 1344–1346.
- Burnston, D. C. (2017). Cognitive penetration and the cognition–perception interface. *Synthese*, 194(9), 3645–3668.
- Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, 8(1), F13–F20.
- Carruthers, P. (2006). *The architecture of the mind*.
- Casasola, M. (2008). The development of infants' spatial categories. *Current Directions in Psychological Science*, 17(1), 21–25.
- Casasola, M., & Bhagwat, J. (2007). Do Novel Words Facilitate 18-Month-Olds' Spatial Categorization? *Children Development*, 78(6), 1818–1829.
- Chow, J., Aimola Davies, A., & Plunkett, K. (2017). Spoken-word recognition in 2-year-olds: The tug of war between phonological and semantic activation. *Journal of Memory and Language*, 93, 104–134.
- Chow, J., Aimola Davies, A. M., Fuentes, L. J., & Plunkett, K. (2016). Backward Semantic Inhibition in Toddlers. *Psychological Science*, 27(10), 1312–1320.

- Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (1994). *A critique of pure vision*. Cambridge, MA, US: The MIT Press.
- Cohen, L. B. (2004). Uses and Misuses of Habituation and Related Preference Paradigms. *Infant and Child Development*, *13*, 349–352.
- Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*(2), 347–382.
- Connolly, K. (2017). Perceptual learning. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2017 ed.). Metaphysics Research Lab, Stanford University.
- Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595.
- Craver, C. F., & Kaplan, D. M. (2018). Are More Details Better? On the Norms of Completeness for Mechanistic Explanations. *The British Journal for the Philosophy of Science*, *0*, 1–33.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153.
- Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *366*(1567), 1149.
- Csibra, G., & Volein, (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, *26*(1), 1–11.

- Cubelli, R., Paolieri, D., Lotto, L., & Job, R. (2011). The effect of grammatical gender on object categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 449–460.
- Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science*, *19*(2), 189–195.
- Duta, M. (2018). *Unpublished images to describe the stimuli in Younger (1986)*.
- Edmiston, P., & Lupyan, G. (2017). Visual interference disrupts visual knowledge. *Journal of Memory and Language*, *92*, 281–292.
- Egan, F. (2018). Function-theoretic explanation and the search for neural mechanisms. *Explanation and Integration in Mind and Brain Science*(1979), 145–163.
- Egyed, K. e. a. (2007). *Understading object-referential attitude in expression in 18-month-olds: the interpretation switching function of ostensive communicative cues*. Poster presented at the Biennial Meeting of the SRDC, Boston.
- Enflo, K. (2020). Measures of Similarity. *Theoria (Sweden)*.
- Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, *2*(10), 704–716.
- Estes, W. K. C. (1994). *Classification and Cognition*. (Oxford Uni ed.). New York.
- Fantz, R. L. (1964). Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Objects. *Science*, *146*(3644), 668–670.

- Ferguson, B., & Waxman, S. R. (2016). Linking Language and Cognition in Infancy. *Journal of Child Language*, 1–26.
- Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development*, 81(2), 472–479.
- Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proceedings of the National Academy of Sciences*, 110(38), 15231–15235.
- Fodor, J. A. (1983). *The modularity of mind: an essay on faculty psychology* (Paperback ed ed.). Cambridge, Mass. ; London: The MIT press.
- Fodor, J. A. (2000). *The mind doesn't work that way : the scope and limits of computational psychology*. Cambridge: The MIT press.
- Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9), 3221.
- Fulkerson, A. L., & Haaf, R. A. (2006). Does object naming aid 12-month-olds' formation of novel object categories? *First Language*, 26(4), 347–361.
- 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.
- Gannon, L. (2002). A critique of evolutionary psychology. *Psychology, Evolution & Gender*, 4(2), 173-218.

- Gergely, G., & Csibra, G. (2013). Natural pedagogy. In *Navigating the social world: What infants, children, and other species can teach us*. (pp. 127–132). New York, NY, US: Oxford University Press.
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 489.
- Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*(5), 677–696.
- Gleitman, L., & Papafragou, A. (2005). Language and thought. In *The cambridge handbook of thinking and reasoning*. Cambridge [etc.]: Cambridge University Press.
- Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64.
- Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, *20*(3), 347–353.
- Glozzi, V., Althaus, N., Mayor, J., & Plunkett, K. (n.d.). *Rethinking the Novelty Preference Task : What does it tell us about Infant Categorisation?*
- Glozzi, V., Althaus, N., Mayor, J., & Plunkett, K. (2013). Primacy/recency effects in infant categorisation. In N. S. M. Knauff M. Pauen & I. Wachsmuth (Eds.), (p. 2410-2415). Austin, TX: Cognitive Science Society. (ID: unige:29491)

- Gliozzi, V., Mayor, J., Hu, J. F., & Plunkett, K. (2009). Labels as features (not names) for infant categorization: A neurocomputational approach. *Cognitive Science*, *33*(4), 709–738.
- Goldstone, R., & Hendrickson, A. (2009, 01). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 69 - 78.
- Goldstone, R., Lippa, Y., & Shiffrin, R. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43.
- Graham, S. A., Booth, A. E., & Waxman, S. R. (2012). Words are not merely features: Only consistently applied nouns guide 4-year-olds' inferences about object categories. *Language Learning and Development*, *8*(2), 136–145.
- Graham, S. A., Keates, J., Vukatana, E., & Khu, M. (2013). Distinct labels attenuate 15-month-olds' attention to shape in an inductive inference task. *Frontiers in Psychology*, *3*(JAN), 1–8.
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357 - 364.
- Gross, S. (2017). Cognitive penetration and attention. *Frontiers in Psychology*, *8*(FEB).
- Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., & Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings: Biological Sciences*, *275*(1653), 2803–2811.
- Gumperz, J. J., & Levinson, S. C. (1991). Rethinking linguistic relativity. *Current Anthropology*, *32*(5), 613–623.

- Haaf, R. A., Fulkerson, A. L., Jablonski, B. J., Hupp, J. M., Shull, S. S., & Pescara-Kovach, L. (2003). Object recognition and attention to object components by preschool children and 4-month-old infants. *Journal of Experimental Child Psychology*, *86*(2), 108–123.
- Han, S., He, X., Yund, E., & Woods, D. L. (2001). Attentional selection in the processing of hierarchical patterns: an erp study. *Biological Psychology*, *56*(2), 113–130.
- Hu, J. F. (2008). *The impact of labelling on categorisation processes in infancy* (doctoral thesis). Department of Experimental Psychology, Oxford University.
- Jouravlev, O., Taikh, A., & Jared, D. (2018). *Effects of lexical ambiguity on perception: A test of the label feedback hypothesis using a visual oddball paradigm*. (Vol. 44) (No. 12). Jouravlev, Olessia: Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada, K1S 5B6, olessia.jouravlev@cunet.carleton.ca: American Psychological Association.
- Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*(3), 339–373.
- Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, *78*(4), 601–627.
- Keates, J., & Graham, S. A. (2008). Category markers or attributes: Why do labels guide infants' inductive inferences? *Psychological Science*, *19*(12), 1287–1293.

- Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805 - 825. (Language and the Motor System)
- Kleene, S. C. (1952). *Introduction to metamathematics*. Amsterdam Groningen: North-Holland P. Noordhoff.
- Koerner, E. F. K. (1992). The sapir-whorf hypothesis: A preliminary history and a bibliographical essay. *Journal of Linguistic Anthropology*, *2*(2), 173–198.
- Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in v1/v2 contributes to categorization of natural scenes. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *31*(7), 2488.
- Koterba, E. A., & Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants's attention and object exploration. *Infant Behavior and Development*, *32*(4), 437–444.
- Ković, V., Plunkett, K., & Westermann, G. (2010). A unitary account of conceptual representations of animate/inanimate categories. *Psihologija*, *43*(2), 155–165.
- Kranjec, A., Lupyan, G., & Chatterjee, A. (2014). Categorical biases in perceiving spatial relations. *PLoS ONE*, *9*(5).
- Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, *65*(2), 145–168.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge: Proceedings of the international colloquium in the philosophy of science, london, 1965* (Vol. 4, p. 91–196). Cambridge University Press.

- Landau, B., & Shipley, E. (2001). Labelling patterns and object naming. *Developmental Science*, *4*(1), 109–118.
- Lawrence, D. (1950). Acquired distinctiveness of cues: II. selective association in a constant stimulus situation. *Journal of Experimental Psychology*, *2*, 175.
- Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology*, *58*(4), 487–524.
- Lipton, P. (2005). Testing hypotheses: Prediction and prejudice. *Science*, *307*(5707), 219–221. doi: 10.1126/science.1103024
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Reference and attitude in infant pointing. *Journal of Child Language*, *34*(1), 1–20.
- Liu, L. G. (1985). Reasoning counterfactually in chinese: Are there any obstacles? *Cognition*, *21*(3), 239 - 270.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review*, *111*(2), 309–332.
- Lucy, J. (1997). Linguistic relativity. *Annual Review Of Anthropology*, *26*, 291–312.
- Lupyan, G. (2007). *The label feedback hypothesis : Linguistic influences on visual processing* (Unpublished doctoral dissertation). Department of Psychology Carnegie Mellon, University Pittsburgh.
- Lupyan, G. (2012a). Chapter seven - what do words do? toward a theory of language-augmented thought. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, p. 255 - 297). Academic Press.
- Lupyan, G. (2012b). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology*, *3*(54), 1–13.

- Lupyan, G. (2015a). Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems. *Review of Philosophy and Psychology*, 6(4), 547–569.
- Lupyan, G. (2015b). Object knowledge changes visual appearance: Semantic effects on color afterimages. *Acta Psychologica*, 161, 117–130.
- Lupyan, G. (2015c). Reply to Macpherson: Further illustrations of the cognitive penetrability of perception. *Review of Philosophy and Psychology*, 6(4), 585–589.
- Lupyan, G. (2017a). Changing what you see by changing what you know: The role of attention. *Frontiers in Psychology*, 8, 553.
- Lupyan, G. (2017b). The paradox of the universal triangle: Concepts, language, and prototypes. *Quarterly Journal of Experimental Psychology*, 70(3), 389–412.
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not Just for Talking. *Psychological Science*, 18(12), 1077–1083.
- Lupyan, G., & Swingle, D. (2012). Self-directed speech affects visual search performance. *Quarterly Journal of Experimental Psychology*, 65(6), 1068–1085.
- 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), 170–186.
- Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences of the United States of America*, 110(35), 14196–201.
- Machery, E. (2009). *Doing without concepts*. Oxford University Press.

- Machery, E. (2010). Précis of doing without concepts. *Behavioral and Brain Sciences*, 33(2-3), 195–206.
- MacPherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84(1), 24–62.
- Macpherson, F. (2015). Cognitive Penetration and Predictive Coding: A Commentary on Lupyan. *Review of Philosophy and Psychology*, 6(4), 571–584.
- Macpherson, F. (2017). The relationship between cognitive penetration and predictive coding. *Consciousness and Cognition*, 47, 6–16.
- Maher, P. (1988). Prediction, Accommodation, and the Logic of Discovery. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1, 273-285.
- Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, 40(2), 230 - 262.
- Mandelbaum, E. (2017). Associationist theories of thought. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2017 ed.). Metaphysics Research Lab, Stanford University.
- Mandler, J. M. (2007). Some differences between percepts and concepts: The case of the basic level. In *Foundations of mind*. Oxford University Press.
- Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science*, 11(1), 53–59.

- Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological Science, 21*(7), 908–913.
- Marchi, F. (2017). Attention and cognitive penetrability: The epistemic consequences of attention as a form of metacognitive regulation. *Consciousness and Cognition, 47*, 48 - 62. (Cognitive Penetration and Predictive Coding)
- Marconi, D. (1997). *Lexical competence*. Cambridge: MIT press.
- Mareschal, D., & French, R. (2000). Mechanisms of categorization in infancy. *Infancy*(1), 59–76.
- Mather, E., & Plunkett, K. (2011). Same items, different order: Effects of temporal variability on infant categorization. *Cognition, 119*(3), 438–447.
- Mayor, J., & Plunkett, K. (2010). A Neurocomputational Account of Taxonomic Responding and Fast Mapping in Early Word Learning. *Psychological Review, 117*(1), 1–31.
- Mayor, J., & Plunkett, K. (2014). Infant word recognition: Insights from TRACE simulations. *Journal of Memory and Language, 71*.
- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences, 14*(8), 348 - 356.
- McCulloch, W. S., & Pitts, W. (1988). A logical calculus of the ideas immanent in nervous activity. In *Neurocomputing: Foundations of research* (p. 15–27). Cambridge, MA, USA: MIT Press.

- Mc Donough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: why nouns tend to dominate verbs in early word learning. *Developmental science*, *14*(2), 181.
- Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, *35*(1), 113-138.
- Mesulam, M. (2008). Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. *Annals of Neurology*, *64*(4), 367–378.
- Michaelson, E., & Reimer, M. (2019). Reference. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2019 ed.). Metaphysics Research Lab, Stanford University.
- Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, *7*(1), 1–9.
- Montemayor, C., & Haladjian, H. H. (2017). Perception and cognition are largely independent, but still affect each other in systematic ways: Arguments from evolution and the consciousness-attention dissociation. *Frontiers in Psychology*, *8*(JAN), 1–15.
- Munakata, Y., & O'Reilly, R. (2000). *Computational explorations in cognitive neuroscience*. Cambridge: MIT Press.
- Nakao, H., & Andrews, K. (2014). Ready to teach or ready to learn: A critique of the natural pedagogy theory. *Review of Philosophy and Psychology*, *5*(4), 465–483.
- Namy, L. L., & Waxman, S. R. (1998). Words and Gestures: Infants' Interpretations of Different Forms of Symbolic Reference. *Child Development*, *69*(2), 295–308.

- Niemeier, S., & Dirven, R. (2000). *Evidence for linguistic relativity*. Amsterdam: Benjamins.
- Niu, H., Wang, J., Zhao, T., Shu, N., & He, Y. (2012). Revealing topological organization of human brain functional networks with resting-state functional near infrared spectroscopy (brain networks by fnirs). *PLOS ONE*, 7(9), e45771.
- Perszyk, D. R., Ferguson, B., & Waxman, S. R. (2016). Maturation constrains the effect of exposure in linking language and thought : evidence from healthy preterm infants. *Developmental Science*(March), 1–9.
- Perszyk, D. R., & Waxman, S. R. (2016). Listening to the calls of the wild: The role of experience in linking language and cognition in young infants. *Cognition*, 153, 175–181.
- Piccinini, G. (2004). Functionalism, computationalism, and mental states. *Studies in History and Philosophy of Science Part A*, 35(4), 811–833.
- Piccinini, G. (2008). Some neural networks compute, others don't. *Neural Networks*, 21(2-3), 311–321.
- Piccinini, G. (2009). Computationalism in the Philosophy of Mind. *Philosophy Compass*, 4(3), 515–532.
- Piccinini, G. (2010). The Resilience of Computationalism. *Ssrn*, 77(5), 852–861.
- Piccinini, G. (2012). *Computationalism* (No. March).
- Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183(3), 283–311.
- Plunkett, K. (1997). Theories of early language acquisition. *Trends in Cognitive Sciences*, 1(4), 146–153.

- Plunkett, K. (2010). The Role of Auditory Stimuli in Infant Categorization. In L. Oakes, C. Cashon, M. Casasola, & D. Rakison (Eds.), *Infant perception and cognition*. Infant Perception and Cognition.
- Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–681.
- Popper, K. R. (1963). *Conjectures and refutations : the growth of scientific knowledge*. London: Routledge.
- Pulvermüller, F. (2018). Neurobiological mechanisms for semantic feature extraction and conceptual flexibility. *Topics in Cognitive Science*, *10*(3), 590-620.
- Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy*, *70*(19), 699.
- Raftopoulos, A. (2011). Late vision: Processes and epistemic status. *Frontiers in Psychology*, *2*.
- Raftopoulos, A. (2019). *Cognitive penetrability and the epistemic role of perception*. Palgrave Macmillan.
- Rauss, K., & Pourtois, G. (2013a). What is bottom-up and what is top-down in predictive coding. *Frontiers in Psychology*, *4*(MAY), 1–8.
- Rauss, K., & Pourtois, G. (2013b). What is bottom-up and what is top-down in predictive coding? *Frontiers in Psychology*, *4*, 276.
- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(4), 1436–1441.

- Rescorla, M. (2017). The computational theory of mind. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2017 ed.). Metaphysics Research Lab, Stanford University.
- Robbins, P. (2017). Modularity of mind. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2017 ed.). Metaphysics Research Lab, Stanford University.
- Robinson, C. W., Best, C. A., Deng, W. S., & Sloutsky, V. M. (2012). The role of words in cognitive tasks: what, when, and how? *Frontiers in Psychology*, 3(April), 1–8.
- Rogers, T. T., & McClelland, J. L. (2008). Précis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences*, 31(6), 689–749.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382 - 439.
- Ross, L. N. (2015). Dynamical models and explanation in neuroscience. *Philosophy of Science*, 82(1), 32-54.
- Rumelhart, D., Hinton, G., & McClelland, J. (1986, 01). A general framework for parallel distributed processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1.
- Samuelson, L. K., & Bloom, P. (2008). The shape of controversy: what counts as an explanation of development? introduction to the special section. *Developmental Science*, 11(2), 183–184.
- Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, 415(6869), 318–320.

- Siok, T. W., Kay, P., Wang, W. S. Y., Chan, A. H. D., Chen, L., Luke, K.-K., & Hai Tan, L. (2009). Language regions of brain are operative in color perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(20), 8140.
- Slobin, D. (2006). From "Thought and Language" to "Language for Speaking". *Rethinking Linguistic Relativity*(July), 70–96.
- Slobin, D. I. (1987). Thinking for Speaking. *Annual Meeting of the Berkeley Linguistics Society*, *13*(January), 435.
- Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*(6), 246–251.
- Sloutsky, V. M. (2010). From Perceptual Categories to Concepts: What Develops? *Cognitive Science*, *34*, 1244–1286.
- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and Categorization in Young Children : A Similarity-Based Model. *Journal of Experimental Psychology*, *133*(2), 166–188.
- 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., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*, *18*(2), 179–185.
- 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), 1478–1492.

- 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.
- Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. *Cognitive Science, 32*(2), 342–365.
- Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: a dumb attentional mechanism? *Cognition, 60*(2), 143–171.
- Smith, L. B., & Samuelson, L. (2006). An attentional learning account of the shape bias: Reply to cimpian and markman (2005) and booth, waxman, and huang (2005). *Developmental Psychology, 42*(6), 1339–1343.
- Speaks, J. (2019). Theories of meaning. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2019 ed.). Metaphysics Research Lab, Stanford University.
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language, 17*(1-2), 3–23.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*(6640), 381–382.
- Stokes, D. (2018). Attention and the cognitive penetrability of perception. *Australasian Journal of Philosophy, 96*(2), 303-318.
- Stone, H. S. (1971). *Introduction to computer organization and data structures*. USA: McGraw-Hill, Inc.

- Strauss, M. S. (1979). Abstraction of Prototypical Information by Adults and 10-Month-Old Infants. *Journal of Experimental Psychology*, 5(6), 618–632.
- Styles, S. J., Plunkett, K., & Duta, M. D. (2015). Infant VEPs reveal neural correlates of implicit naming: Lateralized differences between lexicalized versus name-unknown pictures. *Neuropsychologia*, 77.
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403–409.
- Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, 10(1), 121–125.
- Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of Human Evolution*, 52(3), 314–320.
- Verspoor, M., & Putz, M. (2000). *Explorations in linguistic relativity*. Amsterdam: Benjamins.
- Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition*, 27(1), 62–75.
- Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition*, 70(3), B35–B50.
- Waxman, S. R., & Booth, A. E. (2001). Seeing Pink Elephants: Fourteen-Month-Olds' Interpretations of Novel Nouns and Adjectives. *Cognitive Psychology*, 43(3), 217–242.
- Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition*, 95(3).

- Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*(May), 258–263.
- Waxman, S. R., & Markow, D. B. (1995). *Words as invitations to form categories: Evidence from 12- to 13-month-old infants* (Vol. 29) (No. 3).
- Werker, J., J.F. Gervain. (2013). Speech perception in infancy: A foundation for language acquisition. In *The oxford handbook of developmental psychology* (Vol. 1: Body and Mind, pp. 1287–1293).
- Werker, J. F., & Fennell, C. (2004). *Listening to Sounds versus Listening to Words: Early Steps in Word Learning*. Cambridge, MA, US: MIT Press.
- Westermann, G., & Mareschal, D. (2004). From parts to wholes: Mechanisms of development in infant visual object processing. *Infancy*, 5(2), 131–151.
- Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 1–10.
- Whorf, B. L. (1940). Science and linguistics. *Technology Review*, 42(6), 229–231.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *PNAS*, 104(19), 7780–7785.
- Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development*, 70(1), 65–77.
- Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences of the United States of America*, 105(36), 13690.

-
- Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, *56*(6), 1574–1583.
- Younger, B. A., & Cohen, L. B. (1986). Developmental Change in Infants' Perception of Correlations among Attributes. *Child Development*, *57*(3), 803–815.