# Labelling and Categorization: Evidence from Experimental Studies on Infants

Mara Floris*

*Abstract:* In recent decades, cognitive psychology has focused on the impacts of linguistic labels on the categorization processes in infants. In this article, the results of two experimental studies, Plunkett *et al.*[1] and Althaus and Westermann[2], are critically analyzed in the context of the experimental literature on this topic. From the analysis, it is possible to identify two effects of language labels on categorization: a "grouping effect" and a "segregation effect". These effects are interpreted within a broader debate on language and thought in which Linguistic Relativity and the debate on the Cognitive Penetrability of Perception are interconnected. Within this framework comes the Language Feedback Hypothesis[3] as a theory that could account for the effects observed experimentally both on infants and adults.

*Keywords:* Language, Categorization, Infants, Linguistic Relativity, Cognitive Penetration

## 1. *Introduction*

The investigation into how language and categorization intertwine in early childhood has intrigued cognitive psychologists and psycholinguists since the late 20th century. With more than twenty experimental studies sharing similar methodologies, this research domain aims to clarify the role of word-image associations in the learning and acquisition of categories. This article delves into these issues through the lens of experimental studies employing

* floris.mara@unisr.it
Università Vita-Salute San Raffaele

1 Kim Plunkett, Joan-Fan Hu, Leslie B. Cohen, *Labels can override perceptual categories in early infancy*, "Cognition", 106, 2, 2008, pp. 665–681.

2 Nadja Althaus, Gert Westermann, *Labels constructively shape object categories in 10-month-old infants*, "Journal of Experimental Child Psychology", 151, 2016, pp. 5-17.

3 Gary Lupyan, *Linguistically modulated perception and cognition: The label-feedback hypothesis*, "Frontiers in psychology", 3, 2012, A54.

the novelty preference task[4], predicated on the observation that children exhibit a greater interest in novel stimuli over familiar ones, thereby allowing inferences about their categorical distinctions based on preference.

The present article adopts a minimalist definition of categorization as the sorting of objects into groups, sidestepping the need to attribute representational content to infant categories beyond their visual properties[5]. This cautious approach addresses, but does not fully engage with, the debated impact of linguistic labels on conceptual development[6]. While the interplay between labels and categorization is acknowledged as potentially foundational to concept learning, this paper does not aim to exhaustively explore this complex relationship. Despite the academic discourse often blurring *concepts* and *categories* this paper maintains a clear distinction.

The concept of 'concept' itself is broadly used yet ambiguously defined in cognitive science. Machery critiques the prevailing definitions and suggests replacing the notion of 'concepts' with more specific terms like 'prototype' 'exemplar', and 'theory', defining concepts as a body of information about something stored in long-term memory and commonly employed in cognitive processes resulting in judgments about that thing[7]. This does not imply, however, that infants categorizing objects necessarily possess 'concepts' of those objects, underscoring that categorization - merely classifying objects based on observable characteristics - does not equate to having stable conceptual categories.

Categorization emerges as a basic cognitive ability, not unique to humans but shared across the animal kingdom, pivotal for understanding the studies at hand. It's crucial to distinguish between conceptual and perceptual categories, a difference emphasized by Mandler[8]. Experimental approaches thus far have focused on infants' responses to the physical aspects of objects, indicating that categorization at this stage is predominantly perceptual. Although this suggests infants might develop perceptual categorization prior to conceptual reasoning, it doesn't rule out their conceptual capabilities. This paper posits that conceptual frameworks may not be essential for explaining infant categorization.

[4] Carmel Houston Price, Satsuki Nakai, *Distinguishing novelty and familiarity effects in infant preference procedures*, "Infant and Child Development: An International Journal of Research and Practice", 13, 4, 2004, pp. 341-348.

[5] Nicholas Shea, *Naturalising representational content*, "Philosophy Compass", 8, 5, 2013, pp. 496-509.

[6] Susan Carey, *The origin of concepts*, "Journal of Cognition and Development", 1, 1, 2000, pp. 37-41.

[7] Eduard Machery, *Concepts are not a natural kind*, "Philosophy of Science", 72, 3, 2005, pp. 444-467; E. Machery, *Doing without concepts*, Oxford University Press, 2009; E. Machery, *Précis of doing without concepts*, "Behavioral and Brain Sciences", 33, 2-3, 2010, pp. 195–206.

[8] Jean M. Mandler, *Foundations of mind. Origins of conceptual thought*, Oxford University Press, 2007, pp. 121-145.

Furthermore, while many psychologists view labelling as a facilitator of categorization, the specifics of this effect are not universally agreed upon. The novelty preference task often used in these studies involves introducing infants to a category of visual stimuli, followed by a test phase presenting a familiar and a new image. Increased attention to the new image suggests recognition of the familiar one. The experimental challenge lies in comparing labelled versus unlabeled stimuli to discern labels' specific effects on categorization, a task complicated by the frequent absence of a "silence control condition". This paper, therefore, focuses on two studies deemed reliable in determining labels' impact on categorization: Plunkett *et al.*[9] and Althaus and Westermann[10], emphasizing the need for direct comparison between labelled and unlabeled stimuli to clarify labels' roles.

## 2. *The Experiments in Plunkett, Hu and Cohen's article*

The stimuli employed by Plunkett et al. were composed of illustrations depicting rudimentary animals. Each animal shares a basic body and facial structure (comprising two circles) but differs in the characteristics of their legs, necks, ears, and tails. Variations include the length of the legs and neck, the thickness of the tail, and the spacing between the ears. Each characteristic could take on one of five possible values (labelled from 1 to 5), making every animal a unique combination of these four numerical values. The study featured two distinct experimental scenarios: the Broad Condition and the Narrow Condition. In the Broad Condition, the combination of features was unrestricted. In contrast, the Narrow Condition imposed correlations between certain traits (for instance, long necks were consistently paired with short legs, and the reverse), as illustrated in Fig. 1.

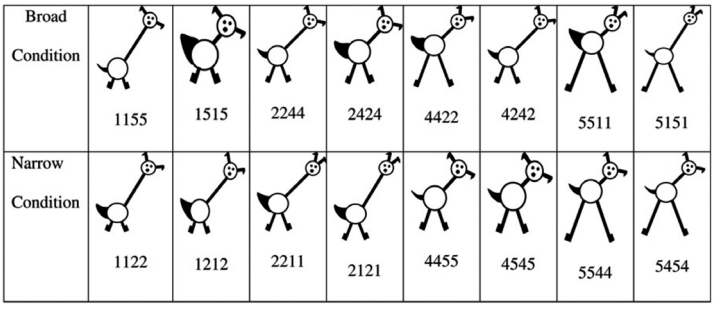



Fig. 1. The Broad Condition and The Narrow condition in Plunkett *et al.* (2008).

9 K. Plunkett et al., *Labels can override*, cit.

10 N. Althaus, G. Westermann, *Labels constructively shape objects*, cit.

The stimuli used in the test phase, which immediately follows the familiarization, are new animals: the overall prototype (3333), the extreme exemplar with low values (1111) and the one with high values (5555), see Fig. 2.

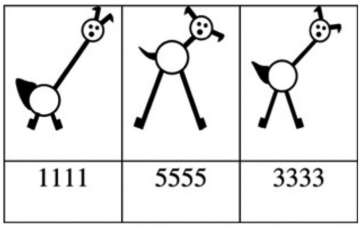



Fig. 2. The test stimuli used by Plunkett *et al.*

They conducted 5 experiments, as reported in Table 1.

| 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 |

In the initial experiment, 24 infants were acquainted with the Broad Condition in silence (mean age 10 months 9 days, range 9 months 21 days to 10 months 12 days). At the testing phase, they exhibited a preference for objects with extreme values (1111/5555) over the median prototype (3333), which serves as the category's prototype. The observation that the extreme examples were recognized, drawing more attention, while the prototype garnered less viewing time, is interpreted as evidence that the objects introduced during the familiarization phase are acknowledged as belonging to the same category. The increased attention to the extreme objects suggests they were perceived as novel. Conversely, in the Narrow Condition, the pattern of preferential looking time was reversed: following exposure to the Narrow Condition stimuli, infants demonstrated longer viewing times for the prototype (3333) rather than the extreme examples (1111/5555). Each extreme example, in terms of geometric similarity, was closer to one of the Narrow Condition's categories and, expectedly, was not the preferred object during testing. Instead, the prototype was deemed more interesting, indicating the recognition of two categories but an uncategorized perceptual space between

them. These initial experiments were pivotal in demonstrating that stimulus manipulation could result in the recognition of one or two categories, contingent on the stimuli with which they were familiarized.

The paper conducted detailed statistical analyses to assess the impact of labelling on infant category formation. During the familiarization phase, a mixed model ANOVA was used to compare the mean looking times of infants across different experimental conditions and blocks. This analysis revealed significant main effects for both the experiment factor ($F(4,115) = 24.886$, $p < .001$) and the block factor ($F(1, 115) = 31.891$, $p < .001$), indicating that the presence of auditory stimuli (labels) increased overall looking time and that infants' attention decreased over time due to habituation. Specifically, infants in the auditory conditions (Experiments 3–5) spent more time looking at the familiarization stimuli compared to those in the silent conditions (Experiments 1 and 2), with significant differences found through post hoc comparisons ($p < .001$).

In the test phase, infants' novelty preferences were analyzed using two-tailed t-tests to determine the proportion of time spent looking at the average stimulus (3333) versus the extreme stimuli (1111 or 5555). Significant preferences were found in several experiments: infants exposed to broad condition stimuli (Experiment 1) preferred the extreme values ($t(23) = -2.99$, $p = .007$), while those in the narrow condition (Experiment 2) preferred the average value ($t(23) = 2.55$, $p = .018$). In Experiment 3, where labels correlated with visual categories, a significant novelty preference for the average stimulus was observed ($t(23) = 2.55$, $p = .018$), mirroring the results of Experiment 2. However, random label assignment in Experiment 4 led to no significant novelty preference ($t(23) = 0.11$, $p = .912$), indicating disrupted category formation. In Experiment 5, the use of a single label resulted in a significant preference for the extreme values ($t(23) = -2.70$, $p = .013$), similar to Experiment 1, demonstrating the label's overriding effect on category formation. Effect sizes (Cohen's d) were reported to quantify the magnitude of differences, and overall, the statistical analyses confirmed that labels could modulate and override perceptual category formation in infants.

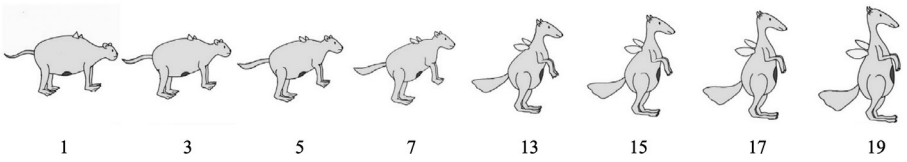
The third experiment replicated the second, but the stimuli were associated with two consistent labels: one subcategory of the Narrow Condition was labelled "rif," and the other "dax." The addition of these auditory cues did not alter the novelty preference task outcome: the subdivision within the Narrow Condition persisted. In contrast, the fourth experiment involved pairing the stimuli with the same two labels from the third experiment in a pseudo-random manner, thereby dissolving any direct association between the subcategories and labels. This misalignment disrupted the previously established categories, with choices at the test phase not deviating from randomness.

The fifth experiment offers a compelling insight into the impact of labels on visual categorization: when stimuli from the Narrow Condition were accompanied by a single label, the outcomes aligned with those from the first experiment, which involved the Broad Condition stimuli. Utilizing one label resulted in the formation of a single category, suggesting that labelling a group of perceptually distinct objects (which in silence are categorized as two separate categories) with the same name leads to them being perceived as members of the same category. This phenomenon, herein referred to as the "grouping effect," aligns with the intuitive notion that if diverse objects are referred to by the same name, they are considered part of the same category.

## 3. *The experiments in Althaus and Westermann*[11]

A contrasting phenomenon is observed in the more recent research conducted by Althaus and Westermann. They developed a set of stimuli that, in the absence of sound, is perceived as a single category but can be divided into two distinct categories when the stimuli are systematically associated with two labels. The stimuli consisted of animal drawings produced using morphing software, resulting in a series of images that blend two distinct animals along a continuum. Rather than manipulating individual features of these animals, the variations between each image were holistic, as illustrated in Fig. 3.

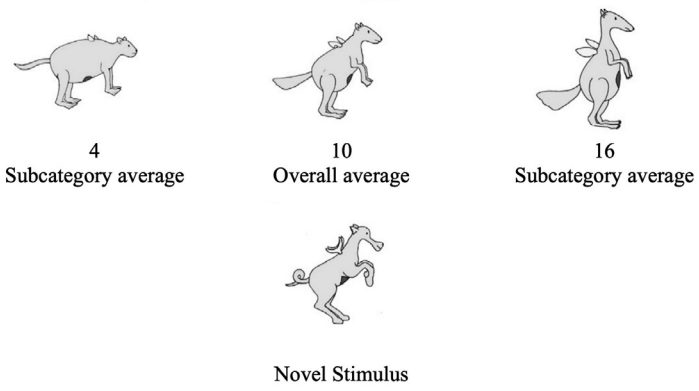
Fig. 3. The stimuli used in the familiarization phase by Althaus and Westermann.




Similar to the Narrow Condition employed by Plunkett and colleagues, Althaus and Westermann's stimuli could be segmented into two subcategories, achieved by introducing a gap within the continuum of morphed stimuli. Their testing procedure was more intricate than that used by Plunkett and colleagues. To eliminate the possibility of a familiarity effect, they conducted comparisons between the overall prototype, the prototypes of the two subcategories, and then contrasted these with a novel stimulus[15], as depicted in Fig. 4.

[11] N. Althaus, G. Westermann, *Labels constructively shape objects*, cit.

Fig. 4. The test stimuli in Althaus and Westermann




In this study, statistical analysis was employed to understand how labelling affects object categorization in 10-month-old infants. Initially, objects presented without sound were perceived as a single category, with the overall average stimulus recognized as familiar. This perception remained unchanged when a single label was applied to all stimuli. However, introducing two distinct labels for two subcategories resulted in infants perceiving the category as divided, with the overall average stimulus attracting more attention. To assess this, a mixed two-way ANOVA compared average looking times across conditions (silent, one-label, two-label) and blocks (first four vs. last four trials). Significant main effects were observed for condition ($F(2,60) = 15.357$, $p < .001$) and block ($F(1, 60) = 14.563$, $p < .001$), indicating longer looking times in the labelling conditions and a decrease over time, showing familiarity. Additionally, preference scores from test trials were analyzed using a mixed effects ANOVA, revealing that two distinct labels led to infants perceiving the stimuli as belonging to two categories ($F(2,59)= 4.96$, $p = .010$, unlike non-linguistic sounds which had no such effect. This contrasts with findings by Plunkett *et al.*[12], where objects without auditory cues were viewed as a single category, but applying two different labels changed this perception, causing the stimuli to be categorized separately.

12 K. Plunkett, N. Althaus, G. Westermann, *Labels constructively shape objects*, cit.

## 4. *Interpretation of the effect*

Prior to delving into the implications of these findings, as well as comparing and contrasting the two studies with the broader research landscape, it's pertinent to address the limitations of the novelty preference task and the conceptualization of category utilized herein. The novelty preference method hinges on the observation that infants devote more attention to novel items over familiar ones, a phenomenon first documented by Fantz[13]. Despite its widespread application within psychological research, ambiguities surrounding the habituation paradigm persist. For instance, there's evidence suggesting that infants may exhibit a preference for familiarity if the habituation period is inadequate, with younger infants particularly showing a tendency towards familiar stimuli when faced with complex scenarios[14]. Nonetheless, even accepting that the novelty preference observed in the aforementioned studies wasn't merely an experimental artefact and is considered reliable, this doesn't fully address the methodological concerns associated with this approach. The novelty preference task offers an indirect, rather than a direct, measure of categorization, merely indicating when one stimulus is perceived as more novel relative to another, without providing a direct assessment of categorization itself.

It's important to note that in Plunkett's experiments, the stimuli used during testing to assess the two potential subcategories were novel extreme examples. Conversely, in Althaus' experiments, the tested stimuli were the prototypes of the two subcategories. When presented in silence, Plunkett's stimuli appear to divide into two categories: the extreme stimuli seem more familiar than the prototype, which attracts more attention. This suggests a perceptual gap between the two categories. Introducing a label seems to bridge this gap: the prototypical item becomes recognized as familiar, and it's plausible that all items spanning the gap between the two categories are similarly recognized. However, the data do not clarify whether the two extreme examples are categorically included. The novelty preference task, by comparing two objects, indicates only which is perceived as more novel, not definitively categorizing the less novel object. Thus, this task provides only a relative preference between two items, with any conclusions about categorical inclusion or exclusion being indirectly inferred. If an object centrally located within the category is deemed more novel than the extremes, it indicates a gap. Yet, if an extreme object – situated at the category's periphery – is preferred over the prototype, its categorical status remains ambiguous.

[13] R.L. Fantz, *Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Objects*, "Science", 146, 3644, 1964, pp. 668-670.

[14] Leslie B. Cohen, *Uses and Misuses of Habituation and Related Preference Paradigms*, "Infant and Child Development", 13, 4, 2004, pp. 349-352.

The data do not definitively reveal what occurs at the category boundaries.

For the very same reason, when in the silent 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 between them, it is plausible to think that the two subcategory prototypes are included in the category. In the two labels case, the labels make the overall average item appear unfamiliar and, presumably, other intermediate items as well. 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[15]. 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[16]. 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 and Westermann are not naturally segregated into two categories. If we look at the way the stimuli are created, we can notice that, as they morph two images, they have the entire perceptual 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[17].

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 categorize, 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 examples seen during the familiarization phase and the overall prototype, instead, is relatively far from them. This experiment seems to show that prototypical effects on categorization are so strong that even if

[15] Barbara A. Younger, *The segregation of items into categories by ten-month-old infants*, "Child Development", 56, 6, 1985, pp. 1574-1583.

[16] Douglas L. Medin, Edward E. Smith, *Concepts and concept formation*, "Annual Review of Psychology", 35, 1, 1984, pp. 113–138.

[17] N. Althaus, G. Westermann, *Labels constructively shape objects*, cit.

infants do not see items close to the prototype, not only they recognize it as familiar, but it is also considered as more familiar than two items closer to the already seen examples. More research is needed to shed light on these questions: what drives natural categorization?

How could an unseen prototype be more familiar than seen examples?

Concerning the grouping effect, there is a study which seems to show a similar result despite the methodological differences. Landau & Shipley[18] tested three groups (2-year-olds, 3-years-olds and adults) with the same set of stimuli consisting of pictures 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. 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/steb"). They were then exposed to the novel 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 steb. 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 examples. 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, although it is worth noticing that this experiment due to its research design lacks a silent control condition.

## 5. *"Grouping Effect" and a "Segregation Effect"*

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. To my knowledge, in the existing literature, this requirement has not been fulfilled[19]. 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 silent 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

[18] Barbara Landau, Elizabeth Shipley, *Labelling patterns and object naming*, "Developmental Science", 4, 1, 2001, pp. 109–118.

[19] Mara Floris, *Language and categorization in infancy*, Aracne, Roma 2022.

affect categorization leads to a positive conclusion. As reported in Ferguson and Waxman[20], a comparison of existing studies suggests that if there is an effect of labels (specifically count nouns) on categorization, this effect depends on the fact that they are names, not merely sounds. This finding sheds light on the experiments in which there was a comparison between a Label condition and a No-Label condition. If we accept 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 facilitative 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[21]. Experiments comparing sounds and words cannot be considered because of a possible overshadowing effect of sounds[22]. 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 categorization. Their results seem to prove that some forms of categorization 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 recognized 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, for example in Waxman and Markow[23] infants (n= 32; with ages ranging from 9.3 to 20.1 months; mean age: 13.5 months) are familiarized with a set of dinosaurs and 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 recognized without a label is weak. It is plausible to think that it is recognized even in silence, but for some reason, the fish is preferred at test only when there is a label.

[20] Brock Ferguson, Sandra R. Waxman, *Linking language and cognition in infancy*, "Journal of Child Language", 44, 3, 2017, pp. 527-552.

[21] M. Floris, *Language and categorization*, cit. p. 20.

[22] Anne L. Fulkerson, Robert A. Haaf, *Does object naming aid 12-month-olds' formation of novel object categories?*, "First Language", 26, 4, 2006, pp. 347-361; Robert A. Haaf, Anne L. Fulkerson, Brandon J. Jablonski, Julie M. Hupp, Stacey S. Shull, and Lisa Pescara-Kovach, *Object recognition and attention to object components by preschool children and 4-month-old infants*, "Journal of Experimental Child Psychology", 86, 2, 2003, pp. 108-123; Sandra R. Waxman, Dana B. Markow, *Words as invitations to form categories: Evidence from 12-to 13-month-old infants*, *Cognitive Psychology*, 29, 3, 1995, pp. 257-302.

[23] S.R. Waxman, D.B. Markow, *Words as invitations*, cit.

Another possibility is that the stimuli used by Waxman and Markow[24] induced some sort of "superordinate-level" categorization, and the stimuli used by Plunkett *et al.* and Althaus and Westermann kept categorization 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 and Markow's studies, instead, the animals used during the familiarization phase are quite different from each other. 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 no categorization occurred.

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.

## 6. *Linguistic Relativity*

The discussion around how language labels impact children's categorization inevitably leads to a broader debate on the influence of language on thought: the "Sapir-Whorf Hypothesis" or "Linguistic Relativity". This theory suggests that the language we speak shapes our thoughts. The widespread interpretation of this hypothesis indicates that language affects cognitive processes. However, the theory's specifics remain somewhat ambiguous; even the term "Sapir-Whorf Hypothesis" is somewhat of a misnomer. Edward Sapir and Benjamin Lee Whorf never jointly proposed the theory, although Sapir was Whorf's mentor. While Sapir's impact on Whorf's ideas is undeniable, Whorf formulated the hypothesis independently, hence it is sometimes more accurately referred to as the Whorfian hypothesis. Whorf's most famous fragment states that:

[24] *Ivi.*

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 organised by our minds and this means largely by the linguistic systems of our minds. We cut nature up, organise it into concepts, and ascribe significance as we do, largely because we are parties to an agreement to organise 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 organisation 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[25].

Initially, the concept of Linguistic Relativity was widely accepted as a fundamental truth, with many psychologists and sociologists studying it as if it were unquestionable. However, the 1970s saw a growing interest in psychological universalism, which began to cast doubt on the theory's absolute validity. This period marked the decline of the theory's more extreme form, linguistic determinism, which suggests that our entire capacity for thought is limited by our language. Identifying the originator of linguistic determinism is challenging, as it doesn't directly stem from Whorf's writings, and no one has explicitly claimed to have developed this iteration of Linguistic Relativity. The idea that language completely governs our cognitive processes seems unlikely, leading to its eventual dismissal. Nonetheless, the 1990s witnessed a resurgence of interest in Linguistic Relativity, evidenced by an increase in empirical research and significant publications on the subject[26]. Currently, more nuanced versions of the hypothesis continue to undergo examination, leading to theoretical progress in the field. Contemporary research explores various domains to pinpoint where the impacts of Linguistic Relativity might manifest, highlighting the significance of identifying which elements of language influence specific cognitive processes. Most of the existing studies focus on whether language can impact perception, conceptualization, and categorization, which are not the same thing. Among the most common topics there

[25] B.L. Whorf, *Science and linguistics*, "Technology Review", 42, 6, 1940, pp. 229-231, (pp. 213-14).

[26] John J. Gumperz, Stephen C. Levinson, *Rethinking linguistic relativity*, Cambridge University Press, 1996; Susan Niemeier, René Dirven (eds.), *Evidence for linguistic relativity*, Benjamins, Amsterdam 2000; Martin Pütz, Marjolijn Verspoor (eds), *Explorations in linguistic relativity*, Benjamins, Amsterdam 2000.

are studies on colors[27], on the effects of grammatical gender[28], on object perception[29] and motion perception[30]. The studies on Linguistic Relativity are different in 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[31].
2. Some effects are active as long as verbalization 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[32].
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[33].
4. Bilingual speakers can switch from habitual schemes when they change the language[34].

Whorf's examples, along with many other studies, typically focus on the second scenario. However, the case study discussed in this article pertains to both the first and third scenarios. This distinction arises because the subjects of the studies under review are predominantly prelinguistic infants. The experiments conducted, or at least a portion of them, contrast the impact of labels on categorization against categorization conducted in silence. Additionally, these studies investigate whether the introduction of varied labels influences the formation of categories.

[27] Franklin, A., et al. *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, 2008, p. 3221. Regier, T., Kay, P., & Khetarpal, *Color naming reflects optimal partitions of color space*, "Proceedings of the National Academy of Sciences of the United States of America", 104, 4, 2007, pp. 1436-1441. Jonathan Winawer, Nathan Witthoft, Michael C. Frank *et al.*, *Russian blues reveal effects of language on color discrimination*, "Proceedings of the national academy of sciences", 104, 19, 2007, pp. 7780-7785.

[28] Roberto Cubelli, Daniela Paolieri, Lorella Lotto, Reno Job, *The effect of grammatical gender on object categorization*, "Journal of Experimental Psychology: Learning, Memory, and Cognition", 37, 2, 2011, pp. 449-460.

[29] Barbara C. Malt, Steven A. Sloman, Silvia Gennari, *et al.*, *Knowing versus naming: Similarity and the linguistic categorization of artifacts*, "Journal of Memory and Language", 40, 2, 1999, pp. 230-262.

[30] Athanasopoulos, P., et al. (2015). *Two languages, two minds: Flexible cognitive processing driven by language of operation*, "Psychological Science", 26, 4, 2015, pp. 518-526.

[31] J. Winawer *et al.*, *Russian blues*, cit.

[32] Dan Isaac Slobin, *Verbalized events*, *Evidence for linguistic relativity*, 198, 2000, pp. 107-138.

[33] Helen De Cruz, *Is linguistic determinism an empirically testable hypothesis?*, "Logique et Analyse", 52, 208, 2009, pp. 327-341.

[34] Stavroula T. Kousta, David P. Vinson, Gabriella Vigliocco, *Investigating linguistic relativity through bilingualism: the case of grammatical gender*, "Journal of Experimental Psychology: Learning, Memory, and Cognition", 34, 4, 2008, pp. 843-858.

The discussion surrounding Linguistic Relativity intersects with another significant debate on the Cognitive Penetrability of Perception (CPP), which posits that our perceptual experiences can be altered by beliefs, desires, or mental states. This notion is widely contested and scrutinized from both theoretical and empirical perspectives across various levels. Theoretically, CPP presents significant implications for epistemology, suggesting that if higher cognitive functions influence perception, then its reliability as a "truth-preserving source of knowledge of the world" comes into question[35]. Proponents of cognitive impenetrability, such as Carruthers[36], Fodor[37], and Sperber and Wilson[38], argue for the modular nature of perception, asserting that its processes are isolated, thus ensuring its dependability as a source of knowledge. In contrast, those advocating for cognitive penetrability challenge the strict modular view of the mind and argue for a perception-based foundation of knowledge that does not necessarily guarantee the preservation of truth. This debate has led to more nuanced discussions regarding the nature of perception and cognition and the demarcation between them, with some recent research even questioning the existence of such a boundary[39]. A focal point of this debate is whether early vision is susceptible to influence by higher cognitive processes, though the broader inquiry extends to the entirety of perception, with a particular emphasis on vision.

In the context of this article, it is important to consider the potential intersection between Linguistic Relativity and the Cognitive Penetrability of Perception (CPP). Linguistic Relativity posits that language influences cognitive functions, while CPP suggests that perception can be shaped by higher cognitive processes. This intersection may provide a framework for understanding the impact of linguistic labels on categorization. The experiments described herein may lie at this intersection, proposing that naming - a linguistic factor - might influence perceptual categorization. Although current literature is insufficient to fully confirm this hypothesis, it represents a promising area where linguistic labelling could affect perception on a cognitive level.

[35] Petra Vetter, Albert Newen, *Varieties of cognitive penetration in visual perception*, "Consciousness and cognition", 27, 2014, pp. 62-75.

[36] Peter Carruthers, *The architecture of the mind*, Oxford University Press, 2006.

[37] Jerry Alan Fodor, *The modularity of mind: an essay on faculty psychology*, MIT Press, 1983.

[38] Dan Sperber, Deidre Wilson, *Pragmatics, modularity, and mind-reading*, "Mind and Language", 17, 1-2, pp. 3-23.

[39] Jacob Beck, *Marking the Perception–Cognition Boundary: The Criterion of Stimulus-Dependence*, "Australasian Journal of Philosophy", 96, 2, 2018, pp. 319-334; D.C. Burnston, *Cognitive penetration and the cognition–perception interface*, "Synthese", 194, 9, 2017, pp. 3645-3666; Carlos Montemayor, Harry H. Haladjian, *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, 2017, e229991.

## 7. *Label Feedback Hypothesis and concluding remarks*

Throughout the past several decades, extensive research has sought to synthesize both the empirical outcomes and theoretical interpretations concerning how labels influence perception, as highlighted by contributions from Ferguson & Waxman[40], Plunkett *et al.*, Robinson *et al.*[41], and Waxman and Gelman[42]. At the heart of these discussions is the debate over whether the influence exerted by labels operates through a top-down or bottom-up mechanism. Plunkett articulates this distinction by describing the processes as supervisory versus non-supervisory, whereas Waxman and Gelman[43] frame the discussion using the metaphors "child-as-data-analyst" and "child-as-theorist". Despite originating from different theoretical backgrounds, these approaches ultimately address the same critical question: Is the effect of labels driven by higher cognitive functions, or does it fundamentally arise from elementary perceptual processes? Given their linguistic nature, labels are commonly believed to shape cognitive processes, reflecting the broad impact of language on cognition. However, it is also plausible that labels serve as auditory stimuli, facilitating perceptual categorization. This dual role implies that the influence of labels may represent a form of Cognitive Penetration, as well as the integration of labels with visual stimuli into a unified perceptual entity.

Within the array of top-down theories elucidating the role of labels in categorization, the Label Feedback Hypothesis stands out for comprehensive examination. Initially introduced by psychologist Gary Lupyan to delineate language's impact on adult cognition, this hypothesis also offers valuable insights into understanding the influence of labels on children. As Lupyan clarifies, his development of this theory was not limited to the effects of labelling alone, despite its pivotal contribution to the discussion, but emerged from the wider discourse surrounding Linguistic Relativity[44]. As mentioned in the previous paragraph, 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.

Winawer *et al.* demonstrated that Russian speakers differentiate certain shades of blue more quickly than English speakers, attributing this to the Russian language having distinct terms for light blue ("goluboy") and dark

[40] B. Ferguson, S.R. Waxman, *Linking language*, cit.

[41] Christopher W. Robinson, Catherine A. Best, Wei (Sophia) Deng *et al.*, *The role of words in cognitive tasks: what, when, and how?*, "Frontiers in Psychology", 3, A95, 2012, pp. 1-8.

[42] S.R. Waxman, S. Gelman, *Early word-learning entails reference, not merely associations*, "Trends in cognitive sciences", 13, 6, 2009, pp. 258-263.

[43] Ivi.

[44] Gary Lupyan, *Linguistically modulated perception and cognition: The label-feedback hypothesis*, "Frontiers in psychology", 3, A54, 2012, pp. 1-13.

blue ("siniy"), unlike English, which uses "blue" for both shades[45]. This linguistic distinction allows Russian speakers to more rapidly distinguish shades of blue when they span across the "goluboy" and "siniy" categories. However, this advantage aligns with English speakers' performance when the shades compared fall entirely within either category. This linguistic benefit vanishes when Russian speakers engage in tasks requiring them to discriminate colors while simultaneously performing a verbal-interference task, such as silently rehearsing digit strings. This phenomenon, indicative of a transient effect, mirrors findings across various studies exploring the interplay between language and cognition. These outcomes suggest that language shapes the perceptual representation of colors, making colors within the same linguistic category seem more similar to each other. Thus, it appears language can alter perceptual space in a way that reflects Whorfian principles.

The ability of language to influence perception in real-time has led to diverse interpretations. Some psychologists argue that the temporary nature of language's effects suggests it does not fundamentally alter thought processes[46]. This viewpoint hinges on the separation of language and concepts, as well as the distinction between verbal and non-verbal processing. From this dichotomous perspective, the transient impact of language challenges the notion of how it could shape conceptual understanding if the effects are not enduring. This raises questions about the extent to which language and cognition are intertwined if concepts must be irrevocably altered by linguistic influence for the relationship to hold. Lupyan's Label Feedback Hypothesis (LFH) aims to resolve this apparent paradox by proposing that language influences perception by "manipulating ongoing perceptual processing online"[47]. This modulation is swift, automatic, and operates within a distributed interactive system. Within the framework of LFH, the act of labelling plays a crucial role in elucidating how language impacts thought, as labels selectively enhance the perceptual features indicative of the labelled category. Lupyan advocates for the view that categorization is a process where "different (i.e., non-identical) stimuli come to be represented as identical in some respect"[48], thereby positioning himself as a proponent of the cognitive penetrability of perception. In exploring how labels influence categorization, scholars divide into two primary theoretical camps: one positing that labels modify percep-

45 J. Winawer, *Russian blues*, cit.

46 Dessalegn, Barbara Landau, *More than meets the eye: The role of language in binding and maintaining feature conjunctions*, "Psychological Science", 19 (2), 2008, pp. 189-195; Lila Gleitman, Anna Papafragou, *Language and thought*, in Keith J. Holyoak, Robert G. Morrison (eds.), *Cambridge handbook of thinking and reasoning*, Cambridge University Press, 2005, pp. 633-661.

47 G. Lupyan, *Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems*, "Review of Philosophy and Psychology", 6, 4, pp. 547-569.

48 G. Lupyan, *Linguistically modulated perception*, cit., p. 4.

tion directly, and another suggesting that labels impact categorization as a higher-level cognitive process, distinct from perceptual processes. Within the faction that views labelling as a form of perceptual modification—or cognitive penetration - opinions diverge on whether this influence affects early or late stages of visual processing, as discussed comprehensively by Raftopoulos[49]. Lupyan aligns with the perspective that perception is susceptible to influence at all stages, advocating for a more radical viewpoint that blurs the distinction between perception and cognition entirely. Echoing Goldstone and Hendrickson[50], Lupyan asserts that the body of empirical data sufficiently demonstrates that perception undergoes alteration: learning to categorize distorts perception and changes certain areas within the perceptual space. According to this view, categorization of objects transcends mere decision-making; it constitutes a fundamental perceptual process. The act of naming, in essence, is an act of categorization, and the process of learning to link labels with objects is considered a form of category training[51]. According to Lupyan:

> The label-feedback hypothesis proposes that language produces transient modulation of ongoing perceptual (and higher-level) processing. In the case of colour, this means that after learning that certain colours are called "green", the perceptual representations activated by a green-coloured 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 colours 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 – upregulated – 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" processing[52].

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 Label Feedback Hypothesis (LFH) challenges the traditional model of conceptual representations that distinguishes between semantic and visual representations, advocating for a view where language-activated representa-

[49] Athanassios Raftopoulos, *Cognitive penetrability and the epistemic role of perception*, Palgrave Macmillan, London 2019.

[50] Robert Goldstone, Andrew Hendrickson, *Categorical perception*, "Wiley Interdisciplinary Reviews: Cognitive Science", 1, 2009, pp. 69-78.

[51] R. Goldstone, Y. Lippa, R.M. Shiffrin, *Altering object representations through category learning*, "Cognition", 78, 1, 2001, pp. 27-43.

[52] G. Lupyan, *Linguistically modulated perception*, cit., p. 4.

tions are multimodal. According to this approach, concepts are represented not through a single modality but through the activation of all relevant modalities; for instance, the visual components of concepts are represented by some of the same neural mechanisms that process their visual perception[53]. Lupyan and Thompson-Schill provide empirical support for this model of hybrid representation. In their research, verbal cues, such as the word "cat", were shown to facilitate a picture verification task more effectively than non-verbal cues like the sound of a cat meowing or a related verbal cue that did not directly denote the object, such as the word "meowing". This efficiency of conceptual activation by language suggests that conceptual representations are more readily activated by linguistic cues. Their results challenge the notion that labels merely facilitate access to non-verbal concepts, as the same conceptual content would presumably be accessible through other cues if this were the case. Instead, language appears to foster distinct types of conceptualizations that are utilized in item categorization[54].

In Lupyan *et al.*, a study similar to those discussed in this article but conducted with adults, two experiments were designed to explore how labels influence the formation of new categories[55]. In the first experiment, participants were asked to discriminate between two types of aliens on a newly discovered planet - those to approach and those to avoid - based on one-by-one presentations. Decisions were met with immediate feedback: a buzz for incorrect categorizations and a bell for correct ones. Participants were divided into two groups, Label and No-Label; in the Label group, after receiving feedback, a printed label (either "leebish" or "grecious") appeared beside the alien. The test phase involved categorizing an alien as either approachable or avoidable, including aliens not presented during the training phase. The second experiment mirrored the first, with the distinction that labels were presented auditorily. An additional Location condition was introduced to rule out the facilitative effect of non-linguistic feature associations. The findings indicated faster categorization with label presence; non-verbal feature association did not yield similar improvements. Moreover, categories formed with verbal cues proved more durable, maintaining their advantage even when labels were absent and in subsequent experiments. These results lend support to the notion that labels enhance categorization performance, highlighting

[53] Lawrence W. Barsalou, *Grounded cognition*, "Annual Review of Psychology", 59, 1, 2008, pp. 617-645; Friedemann Pulvermüller, *Neurobiological mechanisms for semantic feature extraction and conceptual flexibility*, "Topics in Cognitive Science", 10, 3, 2018, pp. 590-620.

[54] G. Lupyan, Sharon L. Thompson-Schill, *The evocative power of words: Activation of concepts by verbal and nonverbal means*, "Journal of Experimental Psychology: General", 141, 1, 2012, pp. 170-186.

[55] G. Lupyan, David H. Rakison, James L. McClelland, *Language is not just for talking: redundant labels facilitate learning of novel categories*, "Psychological Science", 18, 12, 2007, pp. 1077-1083.

the enduring positive impact of labels, a contrast to Plunkett *et al.*

Linking these findings to the Label Feedback Hypothesis (LFH) suggests the necessity of auxiliary hypotheses, such as the idea that categorization processes in young infants and adults share common principles. However, current research suggests differences in how infants and adults categorize and name categories, raising questions about the comparability of these processes across age groups[56]. Further investigation is required to determine the significance of these differences. Additionally, examining whether infants actually learn labels during such experiments and, if not, whether a feedback mechanism operates at the neural level in the absence of a learned label is crucial for understanding the breadth of label effects on categorization.

[56] Eef Ameel, Barbara Malt, Gert Storms, *Object naming and later lexical development: From baby bottle to beer bottle*, "Journal of Memory and Language", 58, 2, 2008, pp. 262-285.