Labeling, Word Mapping and Categorization in Monolingual and Bilingual Infants

Alexandra Ibáñez Lillo

TESI DOCTORAL UPF/ 2016

DIRECTORS DE LA TESI:

Dr. Albert Costa
Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra

Dr. Luca Bonatti
Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra

Universitat Pompeu Fabra
Barcelona

Acknowledgements

It is often said that a dissertation is an individual work but one does not need to think much to rapidly realize that this is not so. For this reason, I would like to thank all of you that in one way or another contributed your grain of salt to getting this work through.

First, I would like thank my directors. Albert, for trusting me when I was just an undergraduate student attending his bilingualism class and for introducing me to the fascinating world of cognitive science. And, of course, Luca for his immense patience, for the scientific guidance, and for always pushing me to go a bit further.

To Ferran Pons for being my first scientific Godfather and being an excellent one.

All these years would not have been the same without my traveling companions: Judith, Júlia, Alba, Paola, Jesús, Marc, Camille, Cinta, Anna, and Irene. Thank you so much for all the shared tupperwares, the laughs, the coffees, the scientific advice, and for always being there for me.

I am very grateful to all the members of the CBC and especially to my two scientific families, the SPB group and the RICO group, for the fruitful meetings, rich discussions, and advice: Cristina, Núria, Elena, Ana, and Irene. Special thanks to Candice for helping me with the laboratory work.

I would also like to thank the support staff for doing important work that sometimes goes unnoticed: Xavi, Sílvia, Katia, Cristina, and Florencia.

Thanks to all the families that every week come to our lab with their little ones and altruistically help us to explore how babies' minds work.

Thanks to my family for raising me a curious person, for their comprehension, and absolute support.

Francesc, I do not have enough words to thank you for your unconditional love and support. Thanks for taking care of all the things you know are important to me during my "absence".

The research reported in this dissertation was funded by a FPI grant (BES-2009-017434) and a project grant (PSI2008-01191) awarded by the Spanish Ministry of Economy and Competency.

Abstract

In order to find the referents of words, infants rely on constrains that guide their interpretations. Currently, little is known about how language environment influences referent identification. Comparisons between bilinguals and monolinguals offer a unique window into this interaction.

The aim of this dissertation is to explore how bilinguals and monolinguals acquire words for categories at different ages as well as their use of disambiguation strategies. Specifically, we asked whether bilinguals' early advantage in executive function, together with the need to learn in noisier contexts, may translate into different word learning strategies.

In Experiments 1-4, we tested 8-, 15-, and 19-month-old infants with a fast categorization task. We also explored if recently acquired knowledge can be used to boost the learning of other information.

In Experiments 5-6, we tested 18- and 19-month-old infants with a familiarization/preferential-looking task on their ability to acquire categories. We asked whether different linguistic experiences may lead bilinguals and monolinguals to form different category representations.

Our findings suggest a much more complex picture of the influence of linguistic experience on referent identification than currently acknowledged.

Resum

Per tal de trobar els referents de les paraules, els infants confien en requeriments que guien les seves interpretacions. Actualment, se sap poc sobre com l'entorn lingüístic influencia la identificació de referents. La comparació de bilingües i monolingües ofereix una finestra única per explorar aquesta interacció.

L'objectiu d'aquesta tesi és explorar com bilingües i monolingües adquireixen noms per categories a diferents edats així com l'ús que fan d'estratègies de desambiguació. Específicament, ens preguntem si l'avantatge primerenc en les funcions executives en bilingües junt amb la necessitat d'aprendre en contextos sorollosos, pot traduir-se en estratègies d'aprenentatge de paraules diferents.

Als Experiments 1-4, van testejar infants de 8, 15 i 19 mesos amb una tasca de categorització ràpida. També vam explorar si coneixement recentment adquirit pot ser utilitzar per promoure l'aprenentatge d'altra informació.

Als Experiments 5-6, vam testejar infants de 18 i 19 mesos amb una tasca de familiarització/mirada-preferencial en l'habilitat d'adquirir categories. Ens preguntem si diferències en l'experiència lingüística poden dur bilingües i monolingües a formar diferents representacions de categories.

Els resultats suggereixen una fotografia més complexa sobre la influència de l'experiència lingüística en l'habilitat d'identificar referents del que actualment es reconeix.

Preface:

Words are the bridge between conceptual knowledge and the world. Somehow, we understand that words represent "concepts" and that concepts represent objects and properties in the world. Infants' appreciation of links between words, objects, and object kinds is a crucial milestone towards understanding how they construct their minds.

One way to explore the links between language acquisition and infants' conceptual system is by studying how labelling influences categorization. In recent years, remarkable advances have been achieved on this topic.

However, infants are not simple organisms who all undergo the same experiences and have the same abilities. Even in the restricted domain of language acquisition, the variations in early human experiences are staggering. Surprisingly, so far, only a few studies have explored how the experience of varied linguistic environments affects the acquisition of word-world relationships. Bilingualism is one of the most extreme cases of variability in language experience. Indeed, bilingual and monolingual infants differ in many fundamental aspects of their linguistic experience—starting from phonology, up to any level of the construction of language representations. We asked whether the rapid acquisition of word-world relationships, and in particular, of the relation between labels and concepts, can be also affected by linguistic experience. The direct comparison of bilingual and monolingual word learning strategies is an opportunity to assess the role of variable human experiences.

This dissertation is aimed at exploring how linguistic experience influences two aspects of word learning: the ability to learn new names for categories and the ability to disambiguate the referent of new labels. It will become apparent that this is not a linear kind of research. Many factors, often underestimated in the literature, contribute to the many layers of complexity of the role of linguistic experience in the creation of world-word relationships. They go from the most general, fundamental, conceptual issues of what exactly these relations mean, to the most detailed, nitty-gritty aspects of the methodology of how an experimenter can put them to the test. We focused on some of them, fully conscious that we are very far from encompassing a whole, satisfactory picture of how infants acquire such relations. To our partial comfort, we feel that we are not alone. We feel that most of the literature shares the same limitations in the understanding of 'the big picture'. However, there is no other way to advance on this topic but to delve into its complexities and try to ride its often stormy waves. And so, we give it a try.

Table of contents

List of figures

1. INTRODUCTION

1.1. The relation between language, the conceptual system, and categorization

The aim of this research is to explore the relation between language and categorization, and to study how differences in linguistic experience may affect this process in infants.

Categorization is a core human ability that helps simplify reality by structuring knowledge in manageable pieces. It releases us from having to remember specific items one by one. It allow us to make inferences about properties without having to experience them directly. It allow us to make predictions about novel category members on the basis of previous knowledge.

Such an important ability starts to develop early on. It has been documented in newborns (Turati, Simion, & Zanon, 2003) and very young infants (Bomba, 1984; Quinn, Slater, Brown & Hayes, 2001; Ferry, Hespos, & Waxman, 2013; Ferry, Hespos, & Waxman, 2010). Infants can also categorize on the basis of many types of dimensions. For example, they can categorize by using object orientation (Bomba, 1984), color (Bornstein, Kessen, & Weiskopf, 1976), events (Baillargeon & Wang, 2002), shape (Quinn & Eimas, 1996), or movement (Konishi, Pruden, Golinkoff, & Hirsh-Pasek, 2016).

Although categorization without language is possible and seen in non-humans as well (e.g. Herrnstein & Loveland, 1964; Wasserman, Brooks, & McMurray, 2015), language and categorization are intimately related (for a review, see Gelman, 2009). Words are our bridge between conceptual knowledge and the world. Concepts refer to objects and classes of objects in the world, and words represent concepts in the mind. That is why it is so important that infants are able to link labels to objects in the physical world. Categorization in infants and children has been widely studied as a means to explore infants' conceptual system. To understand what kind of relations between words and referents infants can form might shed light on their initial concepts and how they develop. But how do infants begin to learn these relations?

An intense debate about the origins of such relations is still underway. Discussing in detail the different views is out of the scope of the current dissertation. In a few words, some researchers hold the idea that at young ages infants construct categories bottom-up by means of domain-general mechanisms that compute regularities on the input with no need for higher conceptual processes. Thus, the role of labels, although relevant, is to support categorization by modulating attention. Labels are thus just another feature of the stimuli such as color or shape (Robinson, Best, Deng, & Sloutsky, 2012). Other researchers suggest that infant categorization is heavily influenced by top-down knowledge, and is driven by infants' initial cognitive abilities plus a set of principles that guide learning (Waxman & Gelman, 2009).

In a well-known study, Waxman and Markow (1995) tested 12- to 13-month-olds in an object manipulation task. Infants were first familiarized with objects of a category (e.g. animals) and later tested on their preference for either a novel member of the familiar category or a novel category (e.g. tools). Infants were assigned to one of three conditions. Some infants listened to a noun during the familiarization trials, others listened to an adjective, and the third group was not given a word. Infants in the noun and adjective conditions showed a novelty preference, while those who did not hear a word did not show any preference. The authors concluded that labelling may facilitate categorization by highlighting objects commonalities. They also suggested that the results supported the idea that infants hold an expectation that words refer to object kinds and that these links might initially be broad, given the lack of differences between the word and the adjective conditions. Since then, an extensive body of research has explored the characteristics of the link between language and categories.

Balaban and Waxman (1997) extended the facilitatory effect of words on categorization found by Waxman and Markow (1995) to 9 month-old infants. Some research raised the question of whether these effects were specific to linguistic labels or if any sound would elicit similar effects by causing a general increase of attention towards the stimuli (Roberts & Jacob, 1991; Balaban & Waxman, 1997).

Some years later it was discovered that from birth infants have a preference for language over other, equally complex, stimuli (Vouloumanos & Werker, 2004; Vouloumanos & Werker, 2007). However, it was not known whether that preference could influence other cognitive domains.

Fulkerson and Haaf (2003) showed that 9- and 15-month-old infants created global categories when objects where accompanied by labels, but did not do so when they were accompanied by sequences of tones, sounds made by the experimenter, or no sound at all. In Fulkerson and Waxman (2007), 6- and 12-month-old infants participated in a familiarization/novelty-preference task. Results showed that infants who saw a set of images paired with a word, but not those who saw the same images paired with a sequence of tones, formed a category. This result showed that, even at younger ages, words have a privileged status with respect to other types of stimuli.

In more recent studies, Ferry et al. (2010) replicated Fulkerson and Waxman (2007) in 3- to 4-month-old infants. At an age when infants are not even able to successfully segment speech, words have a facilitatory effect over tones on categorization.

In perspective, evidence suggests that the link between language and categorization seems to follow a process of narrowing, from a broad connection to a more fine-tuned relationship, as proposed in Waxman and Markow (1995). Further confirmation came from Ferry et al. (2013). In these studies, 3-, 4-, and 6-month-old infants were presented with a set of images that could be paired with lemur vocalizations or backward speech. Results showed that at 3 and 4 months of age, non-human vocalization facilitate categorization as much as human speech does (Ferry et al., 2010). Interestingly, non-human vocalizations did not facilitate categorization per se at 6 months of age. Infants who listened to backward speech failed to form a category at all the three ages. Altogether, results suggest that early in development the link between language and categorization is broad, including human and non-human vocalizations, but that over time this relation is tuned only to human speech.

One more aspect that infants consider about the label-category pairing is its consistency. At 12 months of age, infants start to track the specificity of the links between labels and their referents. Thus, variability in labeling induces object individuation, while consistent labeling induces object categorization (Waxman & Braun, 2005). This result suggests that not only does language facilitate categorization, but that it also has a causal role in the formation of the category to be acquired. In the same line of thought, Plunkett, Hu, and Cohen (2008) presented 10-month-old infants with a novelty preference task, familiarizing them with one of the following conditions. In one of them, one label was paired with two close categories; in another condition two labels were consistently paired with two categories; in a third condition the relation between each label and category was perturbed, making the correlation between them random. In a final condition, infants did not listen to any sound. The authors found that when one label was paired with two groups of objects, infants formed one category. When two labels correlated with two visual categories, infants acquired two categories. Uncorrelated labels disrupted categorization. Finally, in the silent condition infants extracted one or two categories depending on the number of categories presented in the familiarization, but, the acquired categories were not different from those acquired in the labelling conditions. Plunkett et al. (2008) concluded that labels had a causal role on category structure, at least in the case of broad category acquisition. Indeed, in it, infants had to override the initial perceptual distinction of the two narrow categories.

In contrast with the lines of research presented above, some authors question the facilitatory effects of labels. Robinson and Sloutsky (2004) found that 8-, 12-, and 16-month-old infants and 4-years-old children show an auditory preference over visual information. In this study infants and children where familiarized with a geometric shape paired to an auditory stimulus (sounds). After the familiarization, infants were shown one more presentation that could be either identical to the previous ones, present a different sound with the same image, present a different shape with the same sound, or change in both dimensions. Infants detected the changes when the sound changed, but failed to do so when it was the visual information than changed. Infants in the control silent condition did notice the visual violation. The authors suggested that infants' and children's auditory preference may have elicited an overshadowing of the visual input. In Robinson and Sloutsky (2007), 8- and 12-month-old infants were presented with a categorization task in a sound, label, or silent condition. Although infants dedicated more attention to the stimuli in the auditory conditions, their performance was higher in the silent condition at both ages. Thus, not only non-linguistic sound, but also linguistic labels, might shadow visual information. The authors suggested that processing information cross-modally may come at a cost and that this can reduce the quality of the encoding for visual information. They also suggested that sometimes overshadowing may indeed facilitate categorization, by reducing the discriminability between category members.

To sum up, in this section we saw that across development infants fine-tune the link between language and categorization and they learn to appreciate new aspects of this complex link. As in other domains such as perception, infants initially respond to a wide range of symbolic elements that may signal categories but with time they fine-tuned the type of input that they will have to process. We ask whether linguistic experience may influence some of these processes and how they are modified across development.

1.2. What is special about growing up bilingual?

Learning a language is a challenging goal. Every person attempting this task as an adult knows how hard it is. However, infants are able to learn with apparent ease and speed. Thus, it is even more surprising to consider the bilingual case. Overall, it is known that bilinguals achieve most linguistic milestones around the same ages as monolinguals (e.g. Sebastian-Galles, 2010; Werker, Byers-Heinlein, & Fennell, 2009), despite the need to learn from noisier input. However, their word learning strategies are not necessarily the same as those of monolinguals.

An extra difficulty when studying bilingualism is that it is a broad construct that includes very diverse profiles. Bilinguals have different ages of acquisition, learn languages of different distances, receive different kinds of input, are of different socioeconomic backgrounds, etc. In this respect, Catalonia is a privileged case study, because the two language have a similar status and there is little risk of incurring into socioeconomic confounding (Morton & Harper, 2007). It is also common to find crib bilinguals—infants who are addressed to in both languages by their parents from birth.

Bilinguals differ from monolinguals in some essential aspects.

The first evident aspect is that bilinguals need to differentiate the two languages. Monolingual and bilingual infants can already differentiate their native language from languages of other rhythmic classes at birth (Mehler et al., 1988; Byers-Heinlein, Burns, &

Werker, 2010). However, newborns cannot discriminate languages that belong to the same rhythmic class (Nazzi, Bertoncini, & Mehler, 1998). Then, when do infants start to discriminate close languages? And, when does this happen for the case of bilingual exposed to Catalan and Spanish? Bosch and Sebastian-Galles (2001) showed that 4.5-month-old bilingual and monolingual infants can discriminate between the two languages. This result provided evidence that initial discrimination abilities might be similar between bilinguals and monolinguals.

Yet, to learn a language, infants rely on more cues than rhythm. At the beginning, infants start discriminating a broad range of phonemes, whether they are present in their language or not. However, across time infants become tuned to their native languages. They lose sensitivity to non-native contrasts and strengthen native perception (Werker & Tees, 1984; Pons, Lewkowicz, Soto-Faraco, & Sebastian-Gallés, 2009). In bilinguals some evidence suggests a delay in discriminating native phonemes (e.g.Bosch & Sebastián-Gallés, 2003). Indeed, it is way after their first birthday that bilinguals converge to their native phonemes, losing discrimination of contrasts that are irrelevant to their languages (Pi, 2015). This is an important achievement because phoneme discrimination is necessary for speech segmentation.

Learning two languages also mean learning a dual lexicon and being exposed from very early on to translation equivalents.

Initially, it was thought that bilinguals may have a unitary lexicon for the two languages (Volterra & Taeschner, 1978). That account predicted an initial stage of confusion in bilinguals in which they would mix the two languages. However, such a confusion is not usually observed and this view is been progressively abandoned. For example, bilinguals and monolinguals show a similar onset for their first words (Petitto et al., 2001; Pearson & Fernandez, 1994). And bilinguals also have similar vocabulary sizes, once both languages are counted (Pearson, Fernandez, & Oller, 1993; Pearson & Fernandez, 1994)

Another difference is that linguistic input is divided between two languages. Thus, bilinguals probably will have less exposure to each of their languages than a monolingual infant has for his/her one language. Less exposure also means less opportunities to grasp the meanings of words.

That is maybe the reason why bilinguals, even if they have similar vocabulary sizes than monolinguals, know less words in each of their languages (Hoff et al., 2012).

Another possible consequence of having less exposure to either language is that bilinguals may have less phonological knowledge. Ramon-Casas, Swingley, Sebastián-Gallés, and Bosch (2009) showed that bilingual toddlers are not sensitive to mispronunciations in familiar words that vary in just one phoneme contrast of their native languages. Instead, monolinguals showed full discrimination at the same age. Similarly, Fennell, Byers-

Heinlein, and Werker (2007) tried to teach 17- and 20-month-old bilingual and monolingual infants new words for objects that differed in just one phoneme. Monolinguals learned the two words at 17 months of age. Bilinguals did not succeed at this task until 3 months later, thus suggesting that phonological development might influence word learning.

Bilingual infants also tend to learn from noisier environments. It is possible that their parents mix both languages when they speak, or that the languages spoken at home and at school are not the same. In Romaine (1995), six types of bilinguals were described and the differences between them were discussed. The patterns ranged from one-parent-one-language (lower variability condition) to two-parents-mixed-languages (higher variability).

To deal with the extra challenges that bilinguals face, infants may have to adapt some of their learning strategies to better suit the needs of their bilingual environment.

Some differences have been found in the way bilinguals and monolinguals respond to the same problem. One example is that bilinguals seem not to employ a disambiguation strategy called Mutual Exclusivity as firmly as monolinguals do. Byers-Heinlein and Werker (2009) argued that the bilingual experience, where more than one name is associated with the same object, may not favor the development of such strategy.

One important adaptation to the linguistic environment might be the enhancement of executive functions in bilingual infants.

The term 'executive functions' (or cognitive control) refers to a large set of cognitive abilities and processes such as planning, working memory, conflict monitoring, inhibitory control, and decision making (Zelazo & Müller, 2002). These abilities start to develop during the first year of life, with important improvements at 4 to 5 years of age. It is not until adolescence that the adult-like stage is reached (Zelazo & Müller, 2002).

One task used to assess executive functions at young ages is the Dimensional Change Card Sorting Task (Zelazo, Frye, & Rapus, 1996). Before the age of 4, children tend to have difficulties performing this task due to the high demand in executive functions (Diamond, 2002). Thus, it offers a good opportunity to explore whether bilingualism has some effect on the development of executive functions. In this task, children are typically asked to sort cards that represent two different dimension (e.g. color and shape) into two different locations. Each location is signaled by a target stimuli. For example, location A can contain “blue” and “square” and location B, “red” and “circle”. During the first part of the task infants are asked to sort the card according to one dimension (e.g. color). Thus, cards showing a blue circle go to location A and cards showing a red square go to location B. At one point of the task, instructions are changed and children are asked to sort according to the other dimension (e.g. shape). Therefore, the locations are now the opposite. A blue circle goes to location B and a red square goes to location A.

When Bialystok and her collaborators tested 4 to 5 year-old bilinguals and monolinguals with this task, they found that bilingual children outperformed monolinguals and had less problems switching to the other sorting dimension (Bialystok, 1999; Bialystok & Martin, 2004).

But, what explains the difference in performance between bilinguals and monolinguals? Bialystok and Martin (2004) hypothesize that its origin might derive from the continuous control that bilinguals have to exert on their two languages. Although the exact mechanisms of bilingual language control are not known, one widespread view is that in bilinguals, the two languages are activated when speaking in either language. Thus, bilinguals would need to inhibit the lexical competitors from the language not-in-use when speaking (e.g., Green, 1998). The mechanisms monitoring the linguistic system could involve domain-general aspects of the executive functions. Therefore, the constant practice of switching and inhibiting languages may translate into a general advantage of executive functioning extending beyond the linguistic domain proper.

Carlson and Meltzoff (2008) tested monolingual and bilingual children on a wide range of tasks that tap into executive functions to try to better characterize the effects found in previous studies. The authors concluded that bilingual children mostly excel on tasks that call for managing conflicting information, such as the inhibition of a prepotent or distracting response.

The bilingual advantage in executive functions has been found across the lifespan. It is maintained throughout adulthood (Costa, Hernández, & Sebastián-Gallés, 2008). It has been suggested that language monitoring may delay the age-related decline of executive functions (Bialystok, Craik, Klein & Viswanathan, 2004; Craik & Bialystok, 2006; Bialystok, Martin, & Viswanathan, 2005). Yet, these studies left unclear when the difference between monolinguals and bilinguals begins to emerge.

Kovács and her collaborators ran seminal studies that began to shed some light on this issue (Kovács & Mehler, 2009a; Kovács & Mehler, 2009a). The authors found that much before the onset of language production, bilingual infants already behave differently from monolinguals, performing better in tasks that involve executive functions. These findings opened up the exciting possibility that, in order to observe an enhancement of executive functions, one need not wait for the language productive stage to begin. Kovács and Mehler (2009a) tested 7-month-old bilingual and monolingual infants with a task in which an auditory or visual cue indicated the location where a reward would appear. Monolingual and bilingual infants showed similar abilities to learn to anticipate those rewards. However, when the location of the appearance of the reward switched, bilinguals were faster at suppressing the old response and reorienting towards the new location. Infants' ability to redirect their attention towards the novel location was, indeed, a measure of inhibitory control.

In another study, Kovács and Mehler (2009b) tested 12-month-old infants on a rule learning task. In the familiarization phase, infants were presented with two auditory regularities, each of which predicted the appearance of a reward in one location of the screen. In the test phase, infants were assessed on their ability to anticipate the reward when presented with novel tokens of each of the regularities. Bilinguals learned the two rules simultaneously, proving to be flexible learners. In contrast, monolingual infants only learned one of the two regularities. The authors suggested that bilinguals' more refined ability to learn structural information may be related to the constant training in sorting the linguistic stimuli in the two languages that the bilingual experience involves, even before infants begin to actually speak these languages.

Thus, preverbal studies on executive functions showed that processing linguistic information coming from two languages may induce changes in the cognitive system, just as it was shown at older ages. These findings were very important in order to understand how a developing mind may be shaped by language. In particular, this is interesting because it might explain how both bilinguals and monolinguals eventually develop similar linguistic abilities despite the fact that bilinguals must learn in more complex, noisier environments.

1.3. The word learning challenge

Have you ever tried to order food in a cafeteria of country whose language you do not speak, and all of a sudden the waiter asks you something whose meaning you cannot immediately guess? Because I know I am in a cafeteria, which is a place where people come to eat, I know he must be referring to something related to food. Also, because he is pointing and looking towards the fridge, I can deduce that he is referring to something that may be inside it. Because the word he repeated more times comes at the end of the sentence, it might be a name. Because I already asked for a sandwich, he might be referring to something in the same taxonomic category. With a few more steps like those, we may end up finding out that he was asking if I wanted a drink.

This example differs in many important ways from the manner in which infants acquire language, but it still highlights the difficult challenge that infants face every day, with apparent ease and speed, when they learn to map the first words to their meanings.

Before learning a word, several tasks need to be accomplished from the moment in which infants listen to a sequence of sounds to the moment in which they can identify the referent of the word.

Infants need some phonological knowledge of their own language/s in order to recognize the sounds of the speech stream and be able to match them to their representation of sound categories (See Gervain & Mehler, 2010, for a review). These sound representations will need to be narrow enough to allow the identification of relevant sounds, but also sufficiently wide to encompass variations in the input such as differences in speakers' voices or slight differences in their utterances.

Because words often come embedded in a speech stream, infants also need to parse the continuous speech signal and find the word boundaries. There is evidence that infants can use statistical learning abilities to process the signal and learn to recognize words by tracking when and where the word sounds are repeated (Saffran, Aslin, & Newport, 1996). Apart from this process, which is essentially statistical, infants are sensitive to many other cues that can help them find words in the speech signal. Some examples are stress patterns (Thiessen & Saffran, 2003), language rhythm (Goyet, Nishibayashi, & Nazzi, 2013), and phonotactics of the language (Sebastián-Gallés & Bosch, 2002). Also, the exaggerated intonation typical of infant-directed speech seems to have some properties that may facilitate segmentation and word learning (Golinkoff & Alioto, 1995; Thiessen, Hill, & Saffran, 2005).

The ability to successfully segment the speech signal has been shown to be relevant for subsequent word learning. Estes, Evans, Alibali, and Saffran (2007) showed that at 17 months of age the successful extraction of word forms from the speech signal facilitated how those word forms are later mapped to objects.

Once infants learn to identify relevant units from the stream, they need to establish links between labels and their referents.

One pre-requisite to mapping words to objects is to understand what a referent can be. Work on object perception shed some light on infants' initial expectations on the nature of referents. It showed that, before the first year, infants seem to understand objects' permanence in time. Infants also seem to understand that objects exist as separate units. From the work of Spelke (1990), we learned that infants seem to hold assumptions that lead them to treat objects as independent units with boundaries that persist in time. Findings from the field of object individuation also helped explore the kinds of cues that infants exploit in order to determine when two objects are identical and when they may constitute different units. At different ages and with different modalities—whose exact details go beyond the scope of the current work—infants can use spatio-temporal relations, object properties, or linguistic information to refer to and individuate objects (Xu, 1999; Xu & Carey, 1996; Xu, Cote, & Baker, 2005)

All these abilities may help infants overcome the paradox that Quine raised: how to discover the referents of labels, given the essential ambiguity of the world and of the relations between words, objects, and categories? This is *the mapping problem (*Quine, 1960): the difficulty of identifying a referent given the little knowledge that the learner may have and the myriad of possible candidates for a word. Yet, infants solve it. How can they use their *a priori* knowledge of object structure in order to learn words? Several mechanisms have been proposed to try to account for their extraordinary word learning abilities.

One widely held view is that infants might possess strong referential expectations, even the first time that they encounter a new word. Carey and Bartlett (1978) first used the term *Fast Mapping* in a study in which they tested 3-year-old children on their ability to reason about a potential referent for a word after a single label presentation (for an alternative account see Smith & Yu, 2008b).

After years of intense research, a complete explanation of how infants identify the proper referent has not been produced. Yet, it is clear that their vocabularies increase fast (Carey, 1978) and that their mappings seem to be rather accurate (Swingley, 2010).

One possibility is that infants and young children are equipped with heuristics (innate or acquired) that in situations of referential ambiguity guide the mapping of the novel label.

Several biases have been identified, acting at different levels. To list some examples among many, infants are biased to understand when they are being addressed (Csibra & Gergely, 2009). Children also show some pragmatic knowledge about how people use language, through the principles of language conventionality and the assumption than words contrast in meaning (Clark, 2016). Representational biases, such as the whole-object principle (Markman, 1991; Markman & Wachtel, 1988), are particularly important for us. This is the assumption that new words are likely to refer to whole objects instead of object parts or properties of the objects. Because words do not refer just to single objects but also to categories, children may also apply a Taxonomic Principle on top of the whole-object constraint. This principle assumes that the new label can refer to other members of the same taxonomic category.

One biase that has received special attention is children's systematic mapping of novel labels to novel objects that do not have a name.

Imagine a situation in which there is a novel label (e.g. Mapu) and two objects in the scene, one for which we already know a name (e.g. car) and one for which we do not know any name. The unknown label will tend to be paired with the unnamed object.

The most accepted explanation of this phenomenon appeals to the mutual exclusivity strategy. *Mutual Exclusivity (ME)* (Markman, 1991; Markman & Wachtel, 1988) assumes that objects can have only one name. Mutual Exclusivity, here understood as a generic term to refer to the strategy of mapping novel names to novel objects, has been extensively documented in children (E.g. Markman & Wachtel, 1988; Halberda, 2006) and more recently in toddlers (E.g. Markman, Wasow, & Hansen, 2003; Halberda, 2003; Mather & Plunkett, 2012; Byers-Heinlein & Werker, 2009). Recently, Saksida (2014), with a very different procedure from those implemented in the literature—which we will return to shortly—found that 4-month-old infants also map novel labels to novel objects. This finding suggests that the disambiguation strategy might not necessarily depend on infant's knowledge of language.

The question of what mechanism is responsible for such an early appearance of a behavior similar to ME remains open, especially at a period in the infant's life which clearly cannot be influenced by a deep knowledge of language and words. We are lacking knowledge on this particular age, but studies on older infants can give us some hints of where to look for some potential candidates.

Mather and Plunkett (2012) conducted a Mutual Exclusivity study on 22-month-old toddlers in which they tested participants on their preference to map a novel label to a novel but previously familiarized object or to a novel object that had never been seen before. They found that infants tended to map the label to the most novel object. This result opens the possibility that infants were guided by an object-novelty effect.

Halberda (2003; 2006) proposed that the Mutual exclusivity strategy could also be solved through a domain general logic mechanism called disjunctive syllogism. This is a Process-of-Elimination in which the known object is rejected as a candidate because it has a name and therefore the selected referent is the novel object. Halberda (2003) tested three groups of 14-, 16- and 17-month-old infants on a disambiguation task where infants listened to a novel label "dax" while they were presented with a familiar picture (e.g. a car) and a novel object (e.g. a phototube). Fourteen-month-old participants showed a preference for the familiar object, 16-month-old infants were at chance and 17-month-olds successfully mapped the novel name to the novel object. This result was interpreted as evidence that the ability to apply a disjunctive syllogism in a linguistic context may develop around 17 months of age.

The picture that the literature paints is complex. There are different degrees of success at very different ages and with very different task demands. In our opinion, probably no single explanation, by itself, can account for infants' behavior. Most probably, infants can adapt their strategies to their experiences and to the different contexts that each experimental design presents.

Particularly relevant for this dissertation is the study of disambiguation strategies in a bilingual context. Different reasons may make mutual exclusivity not particularly useful for bilinguals. For example, bilinguals need to track the language in which the labelling occurs. They may not know the name of the familiar object in the two languages. Or, they may be more willing to accept two names for the same object because they have experience with translation equivalents. A few studies approached this issue.

Houston-Price, Caloghiris, and Raviglione (2010) found that monolingual infants aged 17 to 22 months, but not bilinguals of the same age, demonstrated the ability to apply the ME principle. Byers-Heinlein and Werker (2009) showed that the degree of adherence to it at 18 months depended on the degree of exposure to language variability. Monolinguals showed a strong tendency to disambiguate the meaning of a novel label by ME. Bilinguals only presented a marginal trend, whereas trilinguals showed no sign of disambiguating the referent of a novel label by using ME. These studies suggest that linguistic experience is fundamental to determining which infants will effectively use the principle, especially in multilingual environments.

In a recent article, Kandhadai, Hall, and Werker (2016) used a rather complex design to test 18-month-old infants. In their task, infants listened to a second label that could be either a second name for an object category or a name for one of its properties (color). Monolinguals and bilinguals were systematic in their interpretations, but they gave opposite meanings to the label. Monolinguals mapped the novel label to the property, thus honoring the principle and showing that it could be used for learning object parts or properties. Bilinguals, instead, did not apply the mutual exclusivity principle and mapped the novel label to the category, thus accepting a second name for it. This finding was interpreted as evidence that bilinguals and monolinguals interpret the label differently, because they have different linguistic experiences.

Overall, we saw that children seem to use principles that help them identify the referent of a new word in ambiguous contexts. We also saw that labels may gain weight as infants accumulate linguistic experience and we also saw that the use or the way in which a principle is applied can be modulated by the specific linguistic experiences that infants carry with them when they approach these tasks.

1.4. The present research

The present work is aimed at exploring how bilingual and monolingual infants learn to map new labels to categories. We are also interested in exploring their ability to identify the referent of new labels in a context of referential ambiguity.

We are motivated by previous findings, which revealed an early bilingual advantage on tasks involving executive functions (Kovács & Mehler, 2009a; Kovács & Mehler, 2009b). Likewise, the wide body of research showing strong connections between language and categorization abilities (e.g., Waxman & Markow, 1995; Fulkerson & Waxman, 2007; Ferry et al., 2010) is at the core of our research question.

Specifically, we ask whether linguistic experience, together with the potential differences in executive functions developed at early ages, may lead bilinguals and monolinguals to display different word learning strategies when they map labels onto categories. We will study this possible interplay in a developmental perspective, by focusing on bilinguals' and monolinguals' capacities at different ages.

We believe that part of the difficulty in understanding the particular role of linguistic experience in the construction of a vocabulary for categories may lie in the many different procedural aspects of the studies already present in the literature. Therefore, we will try to study how different learning contexts may create environments in which infants' linguistic experiences play a different role. In a series of experiments presented in Chapter 2, we will test 8-, 15- and 19-month-old bilingual and monolingual infants' ability to rapidly pair a novel label to a category after a few brief presentations. We will also evaluate their capacity to flexibly exploit this new label-category pairing to bootstrap the referent identification of a second label.

In another series of experiments presented in Chapter 3, we will evaluate 18- and 19-month-old bilingual and monolingual infants on the same abilities, studying how they learn names for categories or identify the referent of a novel label in referentially ambiguous contexts, but with a different procedure. Specifically, we will use a procedure which, in our opinion, poses a less demanding weight on executive functions and is more akin to a word learning task. We believe that a developmentally comparative study of how infants approach these different tasks and, in particular, how their linguistic environment may affect them, may help us clarify some of the variations and the many mysteries that still pervade the topic of the acquisition of the word-category relations. The picture that we will come up with may not be as linear as one would like to have it. This is especially true given that infants are complex organisms, infants' experience with language is complex and varied, and the relation between words and objects is complex. Only by continuing to develop cross-comparisons between ages, tasks, and varied linguistic environments, and only by finding ways to describe what infants do at the finest levels of temporal details, can we hope to find some spark of understanding of this fundamental issue in human cognitive development. This thesis has the ambition to contribute, with its strengths and limitations, to this painstaking but exciting intellectual enterprise.

2. EXPERIMENTS 1 TO 4: Infant Contingent Categorization task

2.1. Introduction

In this experimental series we adapted an anticipatory eye movement procedure to test bilingual and monolingual infants on their ability to quickly learn how to label categories at 8, 15, and 19 month of age.

The procedure was created originally by McMurray and Aslin (2004) to assess categorization in preverbal infants and was modified in Saksida (2014) to teach infants two auditory patterns simultaneously.

One of the challenges when working with preverbal infants or with infants at the onset of language production is how to assess their response non-verbally.

In the recent years, there has been an explosion of new methodologies to study word learning in infants (e.g. Bergelson and Swingley, 2012; McMurray & Aslin, 2004; Saksida, 2014). These methodologies look for signs of more proactive behavior on the part of infants. Methodologically, they also present multiple test trials across the experimental session. One of the reasons for this change lies in the growing implementation of eye tracking techniques, and other technological advances in infant laboratories, that increase measurement accuracy.

Following this line, we adapted the anticipatory eye movement task seen in Saksida (2014) into category acquisition task. We will call this the Infant Contingent Categorization Task.

We decided to modify this paradigm for our purposes because there is evidence that it can successfully induce label-object pairs (Saksida, 2014 p. 76) and teach infants two simultaneous auditory regularities (Kovács & Mehler, 2009b).

In the Infant Contingent Categorization task infants were presented with pairs of pictures belonging to visually distinct categories. Arbitrarily, one of the categories is associated with a name and becomes the target category. The other category remains nameless. The aim of task is to asses whether infants can rapidly map a novel label to a visual category referent. The paradigm presents a visual reward for successful orientation towards the target category—for example, a picture moving on the screen. So, the key to succeed in this paradigm is to look at the target category members, which are the only pictures that trigger the reward—the animation of the picture—when the infant gazes at them.

The main strength of this task is that it does not require a long familiarization phase before the test trials. Also, trial length tends to be shorter than in other types of procedures, allowing for repeated presentations of test trials without boring the infant. But more importantly, it reduces the infant's memory load because familiarization and test trials alternate. Therefore, testing does not need to be postponed to the end of the task, reducing the distance between the familiarization phase and test trials. The shorter trial length also decreases the time between encoding and recall as well as the distance between reinforcers. For these reasons, it can be a particularly successful paradigm if applied to younger infants.

Another advantage of the paradigm is that it is contingent on the infant's looking behavior—as determined by the eye tracker—requiring a proactive response from the infant. Since infants can anticipate the location of the reward, the experimenter can test the expectations formed by infants on the basis of previous exposures and not merely observe their reaction at the final resolution of the trial (for a similar claim see Téglás & Bonatti, 2016). This distinction, which may seem subtle, has been shown to make a difference in some tasks. For example, by means of an anticipatory eye movement task, Albareda−Castellot, Pons, and Sebastian-Galles (2011) demonstrated that 8-month-old bilinguals show the same native contrast discrimination abilities observed in monolinguals. Before that work, it was thought that bilinguals had lost their ability to discriminate. However, this conclusion was not independent of the procedure used. Previous to Albareda−Castellot et al. (2011) bilingual infants were tested with measures based on recovery of attention alone.

Categorization is another domain that typically has been measured through Familiarization/Novelty- Preference (e.g. Balaban & Waxman, 1997; Waxman & Braun, 2005; Younger & Cohen, 1985; Plunkett, Hu, & Cohen, 2008). Most currently existing studies present infants with passive exposure, and sometimes contemplate one single evaluation trial. Compared to the procedures based on the recovery of attention, our procedures allow researchers to collect multiple measures of the variables of interest, because they do not depend on a one-shot surprise response.

Following the same line of thought, the task always presents two images simultaneously. Since infants cannot allocate attention to both pictures at the same time, they are forced to make a decision. Thus, looking patterns detected during the task reveal how infants distribute their attention to the competing stimuli in the timespan of the experimental trial.

The presence of simultaneous pictures provides an opportunity to directly compare. It is been suggested that the comparison between members highlights commonalities and differences and may facilitate categorization (Gentner & Medina, 1998)

A final improvement is that it allows one to study the temporal course of infants' behavior across the trials. To allow for this investigation, we continuously recorded infants' gazes across time. This procedure shows infants' responses relative to a time-locked event, such as a label potentially referring to a category. To assess label mapping, it becomes important to know the timing of the effects since the very same response in one or another time window may indicate different processes. For example, effects at the very first test trial may reveal a pretty quick fast-mapping procedure across the experiment, as opposed to effects that appear only at the end of the experiment. Likewise, in each single trial, effects that occur in the vicinities of a label presentation may indicate the presence of an established association between labels and referents. On the other hand, effects that may occur far from a label presentation in a given trial may indicate that infants cannot find a label reference directly but have to reason through the different possibilities, perhaps through a series of gazes at both pictures. This reasoning in order to come to a decision about the relation between sound and meanings may be perhaps by exclusion. Average looking time scores do not allow one to construct such a rich picture, because they do not tell when an effect occurs.

Thus, we combined a proactive categorization procedure with online looking time measurements, so that we could not only inspect the macrostructure of the looking pattern, but also dig into its microstructure.

As we wrote above, the aim of the task is to teach infants to recognize a visual category associated with a label. During the task infants were faced with pictures pairs depicting exemplars of two distinct animal categories: fish and dinosaurs. We chose two animate categories so as to avoid categorization on the basis of the animate/inanimate distinction (Rakison & Poulin-Dubois, 2001; Rostad, Yott, & Poulin-Dubois, 2012). At the same time, we also wanted relatively rare categories, in order to ensure that if infants create categories, this process occurs in the laboratory and is due to the category induction procedure we implemented. Although it is difficult to control for previous exposure, parents tended to say their child had limited experience with this sort of images.

In the familiarization trials, every time a new image pair was shown infants listened to a pseudo-word that had been associated with one of the visual categories. If the infant fixated on the named category member in the time window after the label, he or she could activate a picture animation that acted as a reward. This paradigm is contingent to infants' behavior because a fixation on the target category always precedes the animation of the picture.

The reward accomplishes two important functions. First, it was a reinforcement for the relationship between the visual category and the label. In the presence of two objects the reward signals the correct answer facilitating the establishment of the association between label and visual category. And second, it motivates infants to continue the experiment.

Of course, the first time that an infant faces a trial it is impossible to find out which picture is the correct one to look at in order to get the reward. Therefore, he or she will perform at chance, or following his or her prior preference to one of the images or sides. However, in principle, already at the next pairing of a picture with a sound, and even more so across successive repetitions, infants may realize the contingency between the label and the animation. Infants who establish the pairing will tend to look more towards the target category.

Alternated with the familiarization trials, the paradigm presented the test trials. Test trials were identical to the previous trials except for one crucial detail. Infants saw two new pictures side-by-side and listened to a label. The critical difference was that this time the images did not move, even if they looked at the side of the screen where the named picture was.

The aim of test trials is to assess whether infants can identify the referent of the listened label and tell it apart from the competitor images.

In most experiments of this series, we used two types of test trials, always presented in a sequential order. In the first experiment of the series we only used the first kind of test trial. Briefly, the test trials were as follows.

In the first test type, categorization trials, a new pair of a fish and a dinosaur was accompanied by the same label participants listened to in the familiarization trials. The purpose of the Categorization trials is to test whether labeling helps infants' categorization in fast presentations with few exemplars, extending what Saksida (2014) found for naming individual items to categories.

In the second type of test trials, which we call *Incidental Mutual Exclusivity (IME) Trials* in reference to a well-known similar paradigm. Here, novel pictures of both categories were presented, but they were accompanied by the auditory presentation of a novel label, that never appeared during the familiarization trials. With this change, the test trial becomes similar to a *Mutual Exclusivity (ME)* problem. Infants listen to a novel label while they see a new member of the familiar category that has been named during familiarization, and a new member of the competitor category that has never been named during the task.

"Standard" ME presents a novel label with a novel object, accompanied by a known object that infants can already name. This procedure tests whether infants can use old knowledge to disambiguate the meaning of new labels, by using the underlying principle that one object can only be associated to one word. Some positive evidence of the use of ME in word learning has been found in toddlers (Halberda, 2003). However, little is know about its emergence across development. With a method similar to the one we are implementing, Saksida (2014) found evidence that infants as young as 4 months may be able to apply analogs to ME. She showed that in a fast word learning task, infants preferred to map the novel label to the novel object. However, nothing is currently known about whether ME is a strategy for learning words for individual objects, or a general strategy for learning class names that extends beyond the concrete objects. Our task could help clarify these issues.

Another interest of our IME trials is that it offers an opportunity to test whether bilinguals and monolinguals approach ME situations differently. There is some evidence suggesting that linguistic background may lead to a differential use of this disambiguation strategy during infancy (Byers−Heinlein & Werker, 2009; Houston−Price et al., 2010). We can think of two ways in which linguistic background may be influencing the use of the ME strategy. One has to do with differences in the kind of linguistic input that bilinguals and monolinguals are exposed to. Bilingual infants, used to listen to more than one label for the same referent, may be less inclined to follow the 'one object, one word' strategy. Thus, bilinguals may be more open to accept a second name for the familiar category and may not apply ME as monolinguals do. The second factor has to do with attentional resources. If instead, the need to disengage from the previously induced label pairing in IME is especially taxing on frontal abilities, then bilinguals may actually apply IME earlier than monolinguals and may be able to reorient towards a novel label-category pairing more easily.

Importantly, our paradigm, departs from traditional versions of ME (e.g. Halberda, 2003; Halberda, 2006; Mervis & Bertrand, 1994; Golinkoff, Hirsh-Pasek, Bailey, Leslie, & Wenger, 1992) in some small but relevant features.

One is that the contrasting visual category used in the IME trials is not completely new. It is only new in the sense that infants had never seen that specific picture in any previous trial. However, the category to which the nameless object belongs has already been introduced in previous trials not associated with the novel label. It is the same competitor category from previous presentations. Because there is previous evidence showing that infants can acquire two categories simultaneously (Plunkett et al., 2008) and that infants can categorize on the basis of visual information alone (e.g. Plunkett et al., 2008; Mareschal, Powell, Westermann, & Volein, 2005) our task may show for the first time that infants can use the ME strategy for learning new labels for categories.

The other difference is that both the 'old' and the 'novel' categories are extracted during familiarization. In contrast with 'standard' ME, where one of the two objects is well known to infants and theater is not, in the IME task the contrasting category is novel or old insofar as infants have been able to extract the first category-label pairing from previous presentations. If this were not the case, infants would lack the necessary anchor point to disambiguate the meaning of the novel label. IME trials require an extra step compared to traditional tasks, that is, learning that the first label refers to one of the categories, by previous recent associations.

In our opinion, those modifications add an important component of executive functioning to the task. We test two aspects of infants' cognitive abilities, both important in order to establish categorization. In Familiarization and Categorization trials, we test infants' flexibility in using labels in directing their fast categorization abilities. Instead, in the IME trials we test whether infants can dynamically modify learning by quickly inhibiting a category they just learned, and exploiting this recently acquired information to bootstrap the acquisition of novel information.

For these reasons, we believe that the study of possible differences between bilinguals and monolinguals will offer a particularly informative window into their learning processes and the influence of language experience on them. Also, because in our opinion the introduction of the IME trials increases task demands, a factor to consider is how the enhancement in cognitive resources across ages influences the ability to apply IME. Thus, we will study whether infants change in how they apply IME at different ages. We will start at 8 months, the age at which bilinguals have been shown to be more flexible learners than monolinguals (Kovács & Mehler, 2009a). And we will end at 19 months, the age at which the ability of using ME has been attested in our laboratory (Cesana-Arlotti, 2015). The age at which infants—bilinguals or monolinguals—show a firm ground on IME, our modified version of ME, will tell us whether ME is more akin to rule-learning or to word acquisition.

To sum up, in this experimental section we will try to explore three main questions.

First, we ask if infants can dynamically extract a category-label association. Specifically, we ask whether infants can learn to identify a category during a brief presentation of a sequence of trials accompanied by a linguistic label. To that purpose, we selected to test infants at ages corresponding to the periods preceding and succeeding the onset of word production and the onset of ME. The different ages allow us to asses the role of language knowledge and cognitive resources in fast categorization.

Second, we ask how the potential fast categorization abilities are affected by the linguistic background. Although there is no specific literature on category acquisition in bilinguals, there are studies reporting bilingual advantages on tasks involving an executive function component (see § 1.2 for a discussion on infants development). These advantages have recently been reported even in preverbal infants (Kovács & Mehler, 2009a; Kovács & Mehler, 2009b). Specifically, we ask if the putative bilingual advantage on executive function could translate into a facilitation of categorization, thus revealing a novel cognitive consequence of the bilingual experience.

If bilingualism improves the ability to quickly categorize, by and large around the same age at which bilingual infants exhibit better skills at learning rules (Kovács & Mehler, 2009b), then bilinguals may anticipate a reward, or anticipate it better than monolinguals, induced by the relationship between a category of pictures and a linguistic label.

Third, we ask whether the process of fast categorization is powerful enough for infants to exploit it and bootstrap novel label learning on the fly, possibly by using an analog of ME at the initial stages of language acquisition. Obviously, a fundamental issue here will be the role linguistic experience may play on this bootstrapping. If sudden changes in the direction of the learning task require a more consolidated development of executive functions then bilinguals may be able to bootstrap novel label acquisition better than monolinguals. If instead, ME is a mainly linguistic strategy, then it is not clear that bilinguals may be any better off. For one, they may not follow the 'one label, one object' strategy that seems to guide monolinguals' vocabulary acquisition. Then, it is possible that their smaller lexicons (Hoff et al., 2012) do not offer a sufficiently solid base on which to seek for a known name that can help to disambiguate the referent of an unknown label. In our opinion, this possibility does not concern our procedure given that we present infants with two novel made-up labels that can not be previously known by infants.

The careful analysis of the temporal course of IME trials, accompanied by the longitudinal analysis of when infants solve our IME task, will inform us on the underlying principles guiding the development of vocabulary bootstrapping in category acquisition.

2.2. Experiment 1: Infant contingent Categorization task at 8 months

In Experiment 1, we test a new categorization procedure that requires a proactive response of participants. At the same time, we test whether both bilinguals and monolinguals can extract novel labels for categories, and if so, whether they do it with comparable temporal dynamics. The procedure has been introduced by Saksida (2014), who showed it to be successful in inducing object-label pairs in monolinguals.

With the Infant Contingent Categorization Task, we will evaluate 8-month-old infants of monolingual and bilingual environments on their fast categorization abilities.

Moreover, we will compare the the performance of infants from different linguistic backgrounds at an age in which there is evidence that bilinguals outperform monolinguals in tasks that require flexible learning (Kovács & Mehler, 2009a). We specifically ask whether an advantage in Executive Functions may translate into better category acquisition for the bilingual group in a paradigm that requires a dynamic response.

2.2.1. Materials and methods

2.2.1.1. Participants

Fifty-six full term 8-month-old infants were retained for analysis: 36 Monolinguals (16 girls. Mean age: 7;23, Range: 7;06 – 8;12) and 20 Bilinguals (10 girls. Mean age: 7;21, Range 6;29 – 8;11). All participants were healthy and free from birth complications according to parent reports.

Twenty-six infants (12 monolinguals and 14 bilinguals) were tested but excluded from the analysis because they cried or refused to stay seated (16); they did not meet the inclusion criteria (4) (see §2.2.2.1.); they were fuzzy (2); they moved during the testing session in a way that did not make the recording of their eyes possible (2); the calibration failed (1); or they were born prematurely (1).

The total rejection rate was 32%. We want to highlight that in this experiment, as well as in nearly all experiments we report, most participants were tested after running one or two tasks. None of the previous tasks had any relation with the topics we assess in this experiment, but infants' attentional resources may have been severely taxed by the previous tasks. Thus their performance in our experiments may underestimate their real abilities.

Infants were considered bilinguals if they had listened at two different languages regularly and had at least 30% of exposure to a second language (L2). The linguistic background was estimated by administrating a detailed language questionnaire (Bosch & Sebastian-Galles, 2001) that collects information about the languages across speakers that the infant has been exposed to from birth. Participants' linguistic background was: 17 Catalan monolinguals, 19 Spanish monolinguals and 20 Catalan-Spanish bilinguals.

Participants were recruited by visiting parents at maternity wards in the Barcelona area. Participants' families were later contacted by telephone. Most families were middle-class. Before running the experiment, parental consent was required. Families were given a small gift and a diploma for their participation.

2.2.1.2. Materials

2.2.1.2.1. Stimuli

We created 36 monochromatic drawings on a grey background. Eighteen of them represented fish and 18 represented dinosaurs.

To create our stimuli, we paired an exemplar of each category (a fish and a dinosaur) semi-randomly, so as to form 18 distinct couples of drawings matched in color and approximate sizes (12 familiarization pairs and 6 test pairs).

The images were edited with Gimp 2.8.14 software and later prepared as animated slides with the software Keynote 5.0.5 (Apple iWork ’09). Finally, the slides were exported as QuickTime movies of 400 x 400 pixels, at 24fps compressed in MPEG-4 video format.

The movies presented the images of the animals looming, increasing and decreasing in size up to 20%.

The linguistic stimuli were two CVCV non-words: Mapu and Doti. The labels were phonetically possible words both in Catalan and Spanish. They were stressed on their first syllable. The audio files were recorded by a female Catalan-Spanish bilingual who spoke in an infant-directed style. The audio files were normalized with the 5.3.14 software in order to match them in duration (1030 ms). Praat

Each label was arbitrarily paired with one of the animal categories according to the experimental condition.

2.2.1.2.2. Apparatus

All stimuli were presented with the software PsyScope X (http://psy.cns.sissa.it) and its Eye Tracker extension TobiiPlus (http://psy.cns.sissa.it/RunTimeInfo/Tobii_and_PsyScope.html) running on an Apple Mac Pro Quad Core 2.8 computer. Infants' eye gazes were recorded with a Tobii T60XL eye tracker at a rate of 60 Hz (http://www.tobiipro.com/product-listing/tobii-pro-t60xl/).

The images were displayed on the 24" high-resolution eye tracker monitor. The pairs of pictures were presented in ports 400 x 400 pixel sized ports located side-by-side. The distance between the ports was 300 pixels.

Audio files were played trough loudspeakers hidden behind the screen.

We also recorded the visual responses of infants through a closed-circuit camera placed discretely on top of the eye tracker, using the iMovie '09 software. Recorded images were sent to a second Apple computer (Mac Mini core i5 1.4) placed outside the experimental room and connected to a screen in TV modality, so as to avoid any delay in showing the infant's behavior in real time. This way, the experimenter could monitor the testing session from outside the experimental room.

2.2.1.3. Procedure

Participants were tested in a soundproof room, dimly illuminated and with its walls covered with dark curtains, so as to prevent infants from becoming distracted during the experiment. Infants were seated on their caregivers' lap at approximately 60 cm from an eye-tracking monitor.

Caregivers were instructed not to interact in any way (by talking, embracing, pointing, etc) with their infants during the experiment. They were further instructed to keep their infants at the level of their hips, leaving them free to move, and to reorient them towards the screen after mentally counting up to 5, in case infants moved away from the screen. Parents wore darkened sunglasses. In case sunglasses were a distraction for their infants, they were instructed to hide them and close their eyes during the experiment.

Before the experiment, infants were calibrated using a 5-point procedure run with PsyScope X software. During calibration participants initially saw drawings played on the screen until participants' eyes were at the optimal distance and height (around 60cm from the screen and eyes located at the upper part of the monitor). Immediately after infants saw a spinning ball that appeared at five different locations (first at the center and later in each of the corners of the monitor) for approximately 2 s in each location. The presentations were accompanied by music to capture attention to the points. Every time the participant fixated on a calibration point, the point disappeared and the next one was presented. At the end of the calibration procedure, participants were shown a large picture together with music while the result was calculated. If more than one calibration point was missing the procedure was repeated until getting at least 4 out of 5 acceptable points to ensure good quality data.

The experimenter monitored the session from outside the room and was blind to the experimental condition. She controlled the experiment by manually monitoring infants' eye directions with keypresses, although the data analysis was run only on the data recorded by the eye tracker.

The experiment consisted of 30 trials organized in 6 blocks. Each block had 4 familiarization trials and one test trial (see Figure 1). The total duration of the experiment was about 3.5 minutes.

All participants saw the same visual stimuli, but to control for potential preferences, infants were randomly assigned to one of 4 possible Conditions: 2 labels (Doti or Mapu) x 2 target categories (dinosaur or fish). Importantly, whatever the category–label association was, it was kept constant within the experiment.

Other features different from category membership or label that could induce categorization were controlled across lists: Side of Appearance of the target picture (Left or Right), Head Orientation of the drawings (Inside or Outside) and Presentation Order of the images pairs (Order 1 or Order 2).

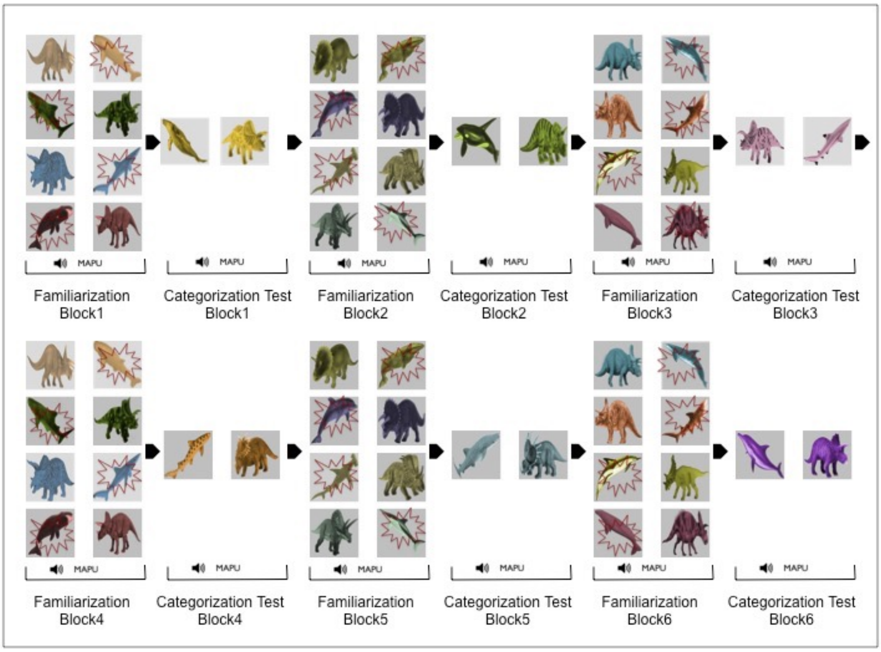


Figure 1. Experiment 1 design. Infants saw a total of 30 trials organized in 6 blocks. Blocks were formed by 4 familiarization trials and one test trials.

The task started immediately after the calibration. Stimuli were presented contingent on infants' oculomotor behavior: infants initiated the presentation of each trial by gazing at the screen.

All trials began with a central attractor (400 x 400 pixels) representing a colorful rotating cross to prevent that infants from orienting towards either target port before the appearance of the images. A fixation on the attractor initiated the trial. The minimum length to consider an interval a fixation was 200ms of uninterrupted looking. After fixating at the attractor, the trial itself started.

Trials were presented in blocks, although infants saw no sign that a block ended and another started. Each block was composed of two kinds of trials (see Figure 1).

Familiarization Trials

Each block presented 4 familiarization trials (or a total of 24 across the experiment). The aim of the familiarization trials was to induce a category by associating it with a particular label.

Then two pictures were presented simultaneously, side-by-side in silence. One of them was always a dinosaur and the other one was always a fish. The pictures varied from one trial to the next. Infants saw each picture only twice across the full experiment.

After 100ms, a label played through the loudspeakers.

After the offset of the label, the response time started. Infants had up to 3500ms in which they could activate the looming of the picture by fixating on the ports where the labeled picture was presented. If infants successfully reached the port within the response time, the video started playing and the picture began looming. Once the movement was triggered, it kept playing uninterruptedly for 2000 ms, after which the trial ended and the next one started. If instead, infants did not orient towards the correct picture within the 3500ms time limit, no picture loomed, a timeout triggered, and the next trial started. (see Figure 2).

Note that the fact that the presentation of the reward was contingent

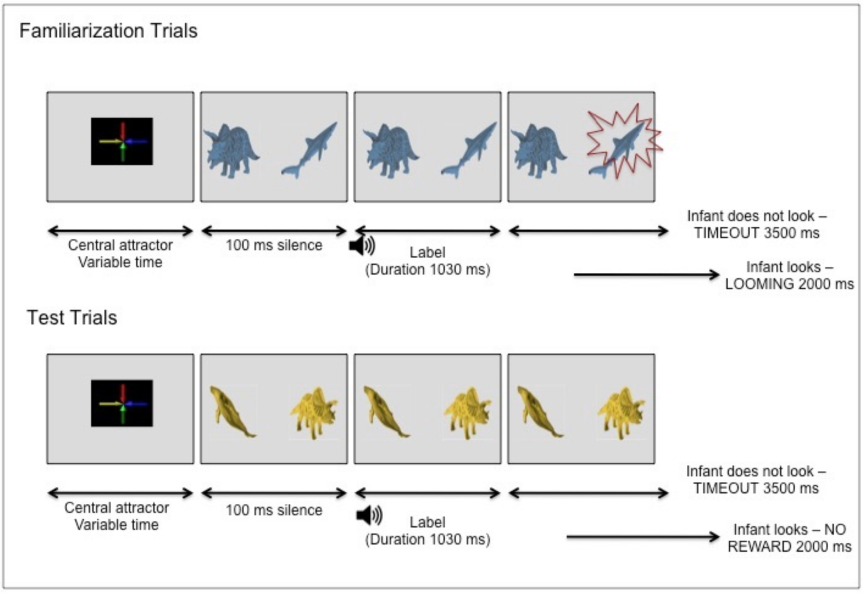


Figure 2. Trial structure of familiarization and test trials in Experiment 1. Both trials were identical, except for the fact that in the test trials the images did not loom.

on infants' oculomotor behavior rendered the duration of the trials variable.

Test Trials

After the four familiarization trials, each block presented a test trial (totaling 6 across the experiment). (see Figure 1). The pairs of images used in the test trials had never appeared during familiarization. Test images were presented only once across the experiment, so as to ensure that infants were generalizing and not simply recalling a previous mapping between a sound and a specific picture.

The structure and timing of the test trials was identical to the familiarization trials, with one important exception. The target picture never loomed, even if infants fixated on the correct image (see Figure 2). That modification permits to study if infants could predict what picture they should fixate in order to trigger the reward.

The fact that the two competitor category exemplars were visible simultaneously also gives us the opportunity of testing any initial preference for one category, besides the referential expectation of participants after listening the label. We can thus also compare infants' change in looking patterns against their own baseline preferences.

2.2.2. Results

2.2.2.1. Scoring

Infants included in the analysis finished the task with at least 70% of time looking at the screen in at least 50% of the familiarization trials and 50% of the test trials. We implemented this inclusion criteria to ensure that infants were engaged in the task.

Four participants were excluded on the basis of these filtering criteria (2 bilinguals and 2 monolinguals).

The eye fixations were sampled by the Eye Tracker at 60Hz. Each datapoint was classified as *Looking To the target*, *Looking To the Non-target* or *Looking to None of the Images*. The target object was defined as the object that had been associated with the label and could loom during the familiarization trials.

To calculate infant's responses, we only considered the gazes that fell into the pre-defined Regions of Interest (ROIs). ROIs coincided with the two experimental ports described in the Procedure section where the Target and Non-Target pictures were presented: a region on the right and one on the left, both 400 x 400 pixels.

The *Proportion of looking time to the target* (PLTT) was used as a measure of Categorization. This was calculated as follows:

$$\textbf{PLTT} = \frac{\text{Data Points to the target port}}{\text{(Data points to the target port + Data points to the non-target port)}} * 100$$

We analyzed the labeling effects by comparing subjects's scores during a B*aseline Window* and a *Post-Label Window*.

We calculated the time course of the PLTT for all familiarization trials collapsed, in order to adjust these time windows to the performance of participants. The details of how the temporal course was obtained and analyzed are described below. The temporal course analysis of the familiarization trials showed that infants, on average, oriented towards the target picture for the first time at around 570 ms after the label presentation in the familiarization trials. This value was defined as the *Orientation Onset*.

The Orientation Onset was used to divide the trial time into two temporal windows of interest:

- **Baseline Window:** It corresponds to the time window that spans from the trial onset to time-point immediately before the Orientation Onset. It also corresponds to the minimal latency infants needed to reach the target port as shown by the familiarization trials. The scores of the baseline period serve as a control for initial preferences not raised by our experimental manipulations.

- **Post-Label Window:** It is the time window in which we examined infants responses to the label. It spans from the orientation onset until 1000ms later. Given that the duration of the trial was variable depending on when the picture started looming, we decided to select a 1 s common fixed time window for all participants to study the infants' responses. This time window ensures some stability in the data, because it corresponds to the window in which most participants contributed eye data[1]. Notice too that this time window does not start immediately after the label, giving infants some time to process the sounds (Swingley & Aslin, 2007).

For each participant, we calculated PLTT during the baseline period and the Post-Label Window. Scores above 50% indicate a

[1] We were forced to do that given that data was noisy at the end of trials due to the progressive drop in data points. The reason is that only infants who triggered the looming late in time would contribute at the end of the trial since the early ones had already finished the looming period.

preference for the category that had been associated with the label in the familiarization trials (familiarity preference). Scores below 50% indicate a novelty preference, in favor of the visual category not reinforced in the familiarization trials. Thus, if infants correctly map the label to the target category, we expect to observe an increase in the PLTT in the Post-Label window.

We also computed the PLTT *across time*, dividing the time period of interest into time bins of approximately 100ms each (6 data points). Bins containing less that 40% of the potential data points were excluded from the analysis. The PLTT across timeBin rejection rate was 38%. measure allows us to study the dynamics of the effects. A consensus has not been reached on how to analyze the temporal time course data in eye tracking studies. Indeed, several criteria can be found in the literature (e.g. Ferguson, Graf, & Waxman, 2014; Borovsky, Elman, & Fernald, 2012). Some of them define success as the presence of an arbitrary number of consecutive time periods for which a *t*-test is significant, without any motivation. Conservatively, we decided to evaluate the temporal course with the Guthrie and Buchwald (1991) procedure, in order to identify the temporal regions in which such effects were maximal. This technique, originally designed for ERP data, uses MonteCarlo simulations to assess how likely it is to observe an effect over a certain amount of consecutive time periods, given a sample of size *N*, *T* time periods and a known data autocorrelation. We note that the results we report would be stronger if we had we chosen some of the less conservative, but more arbitrary, criteria sometimes reported in the literature.

The nature of the time course data allows us to make two types of comparisons. First, we can look for significant changes in the PLTT in response to the labels. We did that by comparing the time points of the Post-Label window against the mean score of the baseline window. Second, we also compared the time courses of bilinguals and monolinguals in order to explore response differences between groups, by comparing their temporal courses at each time-bin.

According to the analysis we chose, for a sample of N= 20-to-30 participants, a window length of 10 time bins (or 1s), assuming the highest degree of data autocorrelation and a significance level of $P < 0.05$, significant t-test comparisons indicate a reportable difference in the looking time pattern. 4 or more consecutive

Given that we expect a familiarity preference in the familiarization and categorization trials and a novelty preference in the IME trials (Experiments 2-4), we used one-tailed *t*-tests in order to compare the experimental conditions against their baseline. However, because we have no firm prediction about the direction of the effect when comparing bilinguals and monolinguals, we used two-tailed *t*-tests for these comparisons.

Finally we also performed a regression analysis in order to explore how infants' linguistic background affects their task performance. For that purpose, we estimated the each infant we estimated the percentage of L2 exposure for each infant. As a measure of success, we extracted a Difference score by subtracting the average PLTT in the Post-Label window to the PLTT during Baseline. Positive Difference score values indicate an increase of looking time to the target, while negative values indicate a reduction of looking time to the target.

2.2.2.2. Results

Preliminary analyses on the Familiarization trials detected no effects of sex (Girls or Boys)($P = 0.6$), or Label (Doti or Mapu)($P = 0.90$). But Target Category did have an effect.

A one-way analysis of variance (ANOVA) on the PLTT in Familiarization Trials, with Target Category (Dinosaurs/Fish) as a between–participants factor and participants as a random factor nested inside Target Category, revealed a significant main effect of Category ($M\ PLTT_{Dinosaurs} = 71.52$, $M\ PLTT_{Fish} = 48.41$; $F(1,56) = 104.79$; $P < 0.0001$). Post hoc Scheffé tests revealed infants had a preference for the category dinosaur over fish.

To correct for baseline preferences for one category, in test trials we compared infant's performance, not against chance, but against their baseline preference scores.

2.2.2.2.1. Familiarization Trials

The results reported here include the averaged responses to all the familiarization trials collapsed.

We ran a two-way mixed-design ANOVA on the PLTT in familiarization trials with time window (Baseline/Post-Label) as a within–participants factor, linguistic background (Bilinguals/ Monolinguals) as a between–participants factor and participants as a random factor nested in linguistic background. The results showed a main effect of Time Window ($M\ PLTT_{Baseline} = 49.76$, $M\ PLTT_{Post\text{-}Label} = 70.40$; $F(1,54) = 167.69$; $P < 0.0001$). There was no effect of linguistic background ($P = 0.84$), nor an interaction between Time Window and Linguistic background ($P = 0.10$). Thus, the results showed that both bilinguals and monolinguals increased their PLTT during the Post-Label period, and that the scores did not differ between groups.

To localize the effects in time, we ran a time-course analysis on the PLTT of the familiarization trials. Both group increased the PLTT above chance. For the monolingual group it was found a time window that started 770 ms after the label offset and lasted 1800 ms (That is from 1900 to 3700 ms after the trial onset). For bilinguals the time window started 670 ms after the label offset and lasted 1600 ms (That is from 1800 to 3400 ms after the trial onset). We did not find significant differences between groups.(see Figure 3).

Thus, the results of the time-course analysis showed that bilinguals and monolinguals significantly increased their PLTT following the label presentation, congruent with previous analyses.

These results suggests that both groups successfully triggered the reward and that once the looming started they stayed focused on the

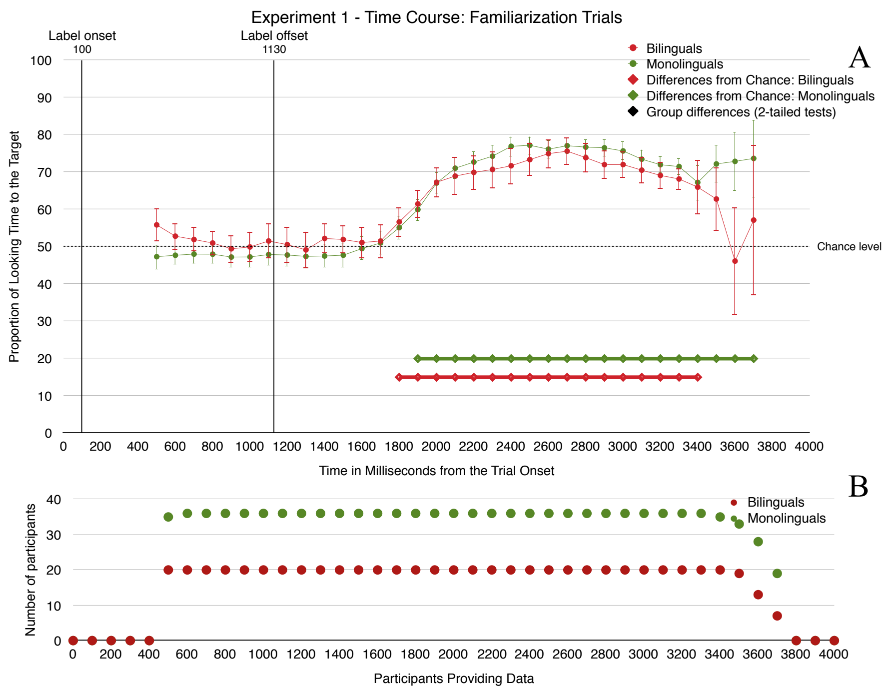


Figure 3. **A.** Symbols representing the PLTT across time in Familiarization Trials of Experiment 1. Vertical lines represent the onset and offset of the label. The Horizontal line indicates the Chance level. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from Chance level for the bilingual and monolingual group and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

target port. It may suggest that infants successfully extracted a category on the basis of the label presentation, or that they extracted it on the basis of visual commonalities between the objects of the categories. Our design does not allow us to determine which hypothesis is correct. However, the late orientation time with respect to the label offset suggests that infants did process the label, indicating that they may have used the linguistic information to direct the way they formed the labeled category.

2.2.2.2.2. Categorization Test Trials

We first obtained averaged responses to the 6 Categorization trials collapsed. For each participant, a Mean PLTT score was calculated for the Baseline Period and the Post-Label period. We ran a two-way mixed ANOVA on the PLTT in the Categorization Trials with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic Background (Bilinguals/Monolinguals) as a between–participants factor and Participants as a random factor nested in Linguistic background. Results did not identify a main effect of Time Window (P = 0.40), or of Linguistic Background (P = 0.66) or an interaction between these factors (P = 0.83) (see Figure 4).

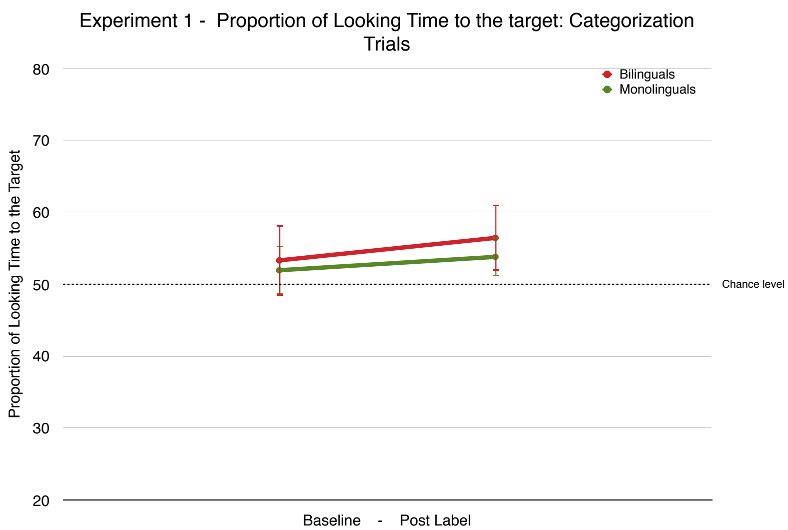


Figure 4. Mean PLTT during the Baseline and Post-Label period for the Categorization Trials of Experiment 1. Bars represent SEs. * indicate effects at $p < 0.05$

Additionally, we conducted the same analysis with Target Category as an independent variable, because it was shown to have an effect on infants PLTT in the previous analysis. We ran a three-way mixed ANOVA on the PLTT in the Categorization Trials with Time Window (Baseline/Post-Label) as a within – participants factor, Linguistic Background (Bilinguals/Monolinguals) and Category (Dinosaur/Fish) as a between – participants factor and Participants as a random factor nested in Linguistic background * Category. Results did not identify any main effect, except for Category as seen in the preliminary analysis. Importantly, there was an interaction between Time Window and Category ($F(1,52) = 7.19$; $P = 0.009$). No other significant interaction was significant. Post hoc Scheffé tests revealed that infants who had been familiarized with the fish category as a target presented a significant increase of PLTT in the Post-Label period ($P = 0.01$) while the ones that were presented with the category dinosaur as a target did not change their PLTT ($P = 0.21$) after the label (see Figure 5).

Putting both analyses together, overall infants had a preference for the dinosaur category. When that category was the target, they maintained their baseline preference in the Post-Label phase. Instead, when the target category was the fish, they overcame this preference and tended to look more at the fish in the Post-Label phase.

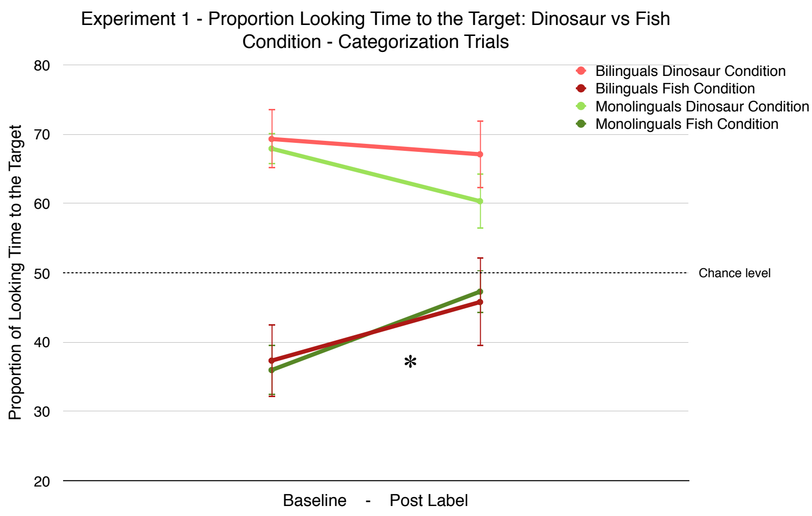


Figure 5. Mean PLTT during the Baseline and Post-Label period for the Categorization Trials of Experiment 1 split by Target Category. Bars represent SEs. * indicate effects at $p < 0.05$

We then computed a time-course analysis on the PLTT in the Categorization Trials. An independent baseline score for the monolingual and the bilingual groups was computed, and each time-bin in the Post-Label window of 1 s was compared against the Mean PLTT score for the respective Baseline (see Figure 6).

We found no sufficiently long temporal stretch in which participants overcame their baseline preferences. There was only a short time window (200 ms) in which bilinguals increased the PLTT. That region coincides approximately with the orientation latencies shown in the familiarization trials, but its significance is unclear. No group differences were found.

Considering the interactions we found in the previous ANOVA, the overall result of the temporal analysis, which collapses all the

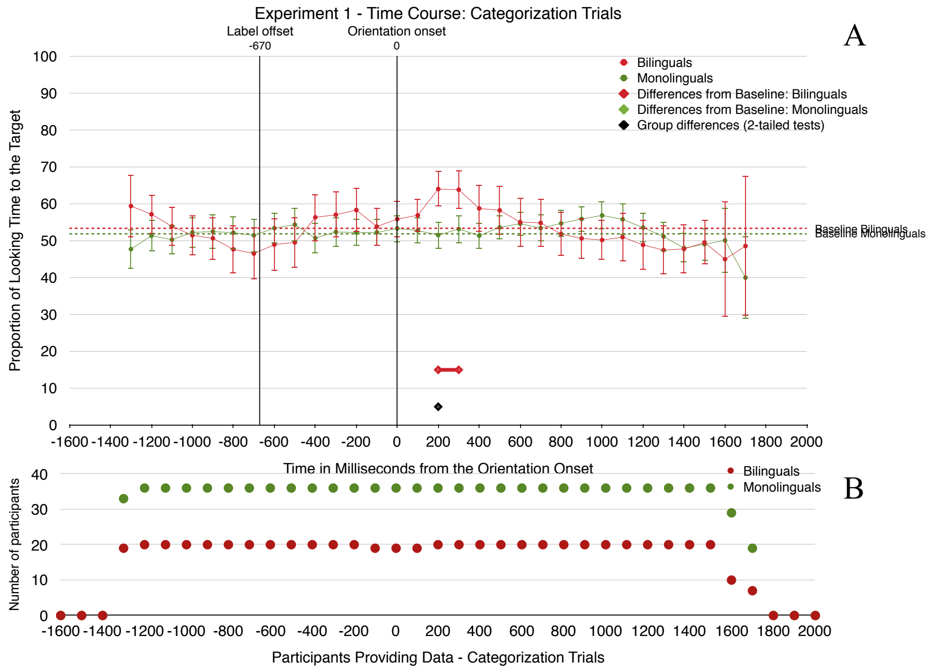


Figure 6. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 1. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p< 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p< 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

categorization trials and averages out which category was the target category, is likely to underestimate potential signs of learning. Thus, we ran two different further analyses: one by splitting the temporal course by target category and a second one by looking at the temporal course of learning across the experiment.

The time-course analysis split by Target Category showed that infants who were familiarized with the “non-attractive” category fish, started the trial looking to the non-target category and increased their PLTT around the orientation onset. Monolinguals presented a long stretch (800 ms) in which they increased the PLTT above their baseline score. The sequence extended from 100 ms to 800 ms after the orientation onset. In bilinguals we found no sufficiently long window (200 ms) around the orientation onset in which they also increased their PLTT (see Figure 7).

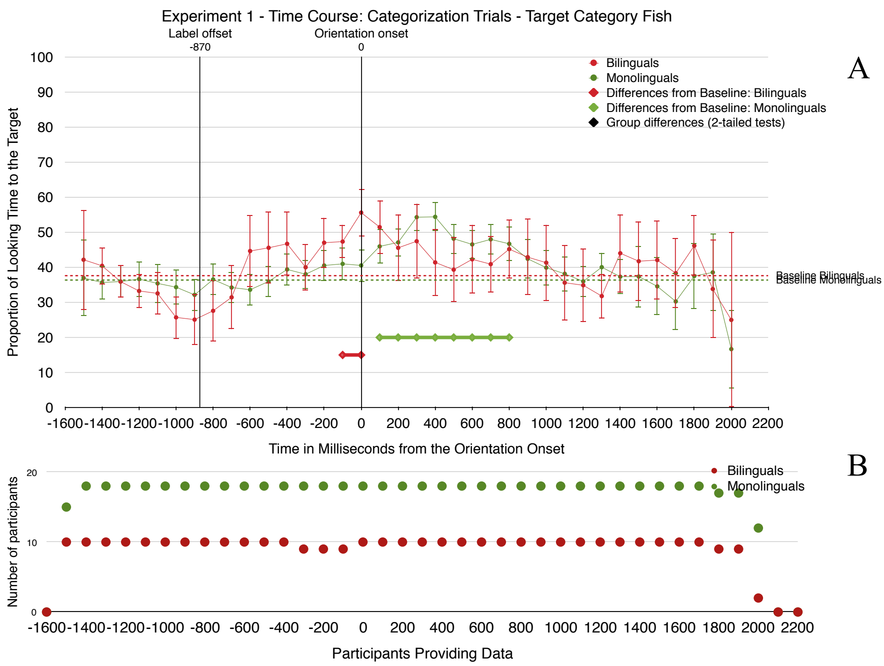


Figure 7. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 1 for infants familiarized with the Target Category Fish. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

The time-course analysis of the infants who were familiarized with the "attractive" category dinosaurs, revealed that they started the trial already oriented to the target category. Overall, they also looked at the stimulus for less time during the trial presentations. The mean PLTT during the Baseline period was above chance in the monolingual (Mean PLTT = 67.92; $t(17) = 7.82$, $P < 0.0001$) and bilingual group (Mean PLTT = 69.30; $t(9) = 4.54$, $P = 0.001$) (see Figure 8).

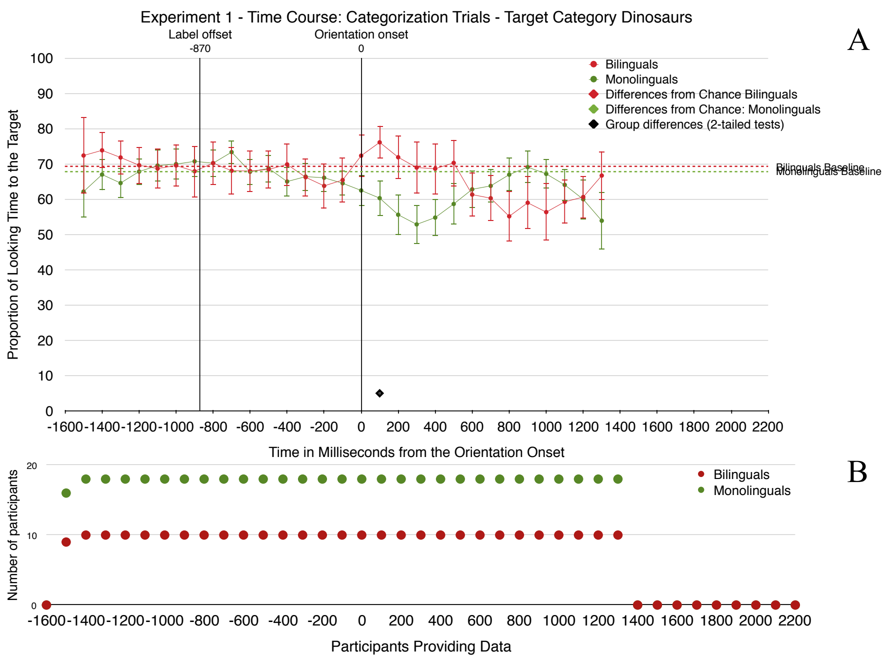


Figure 8. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 1 for infants familiarized with the Target Category Dinosaur. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

As expected, the time-course analysis of the PLTT of the infants who were familiarized with the dinosaur category did not reveal any sequences for bilinguals or monolinguals. Results suggest that monolinguals and bilinguals did not increase their PLTT after the label most probably because they where already oriented towards the target category during the baseline period.

The results split by Target Category suggests that infants overcame their initial preferences and reacted to the label increasing their PLTT if they were not oriented to the target image (Fish category) or they maintained the attention on the Target Category (Dinosaur category) if they were already oriented.

In order to better understand the role of exposure to repeated examples of the pairing between a label and an object of a target category, we ran a restricted time course analyses on the two halves of the experiment separately: Blocks 1 to 3 and Blocks 4 to 6.

Results revealed that, in the first half, monolinguals were very clearly oriented towards the target category. A clear temporal region between 500 ms and 1000 ms from the onset of the orientation period appeared in which monolinguals looked more at the target category than in their baseline. Bilinguals tended to do so at an earlier time window (between 200 ms and 300 ms after the orientation onset) but in too short of a sequence (see Figure 9).

Even if both groups increased their PLTT during the Post-Label window, suggesting they correctly identified the target, we also found marginal group differences between bilinguals and monolinguals from 900 ms to 1100 ms after the orientation onset. The finding suggests that the dynamics of the response developed differently in the two groups. Monolinguals showed a slower, but more pronounced, response than bilinguals.

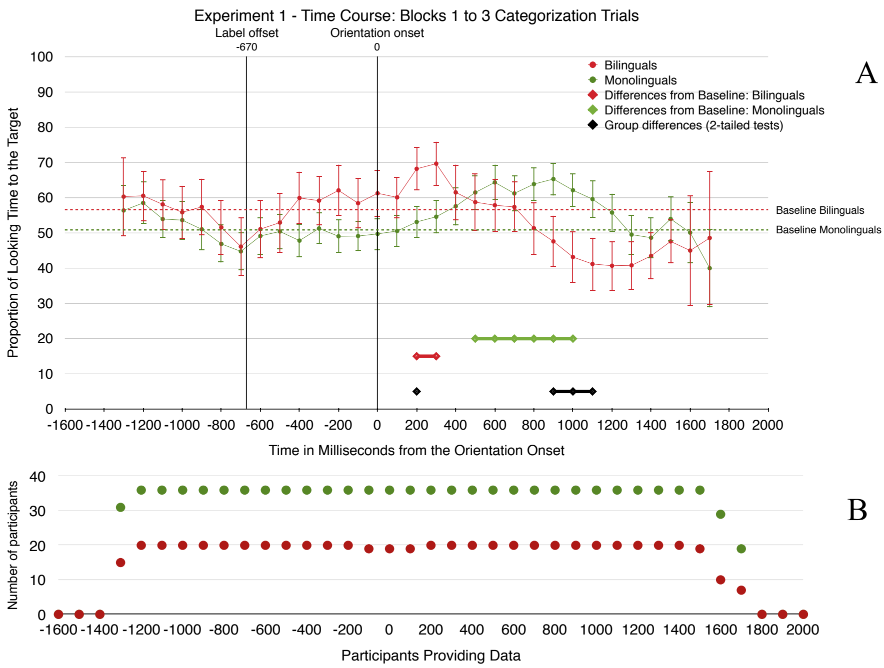


Figure 9. **A.** Symbols representing the PLTT across time for the first half of Categorization Trials of Experiment 1. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

In contrast, in the second half of the experiment the analysis did not identified any effect. This might point to a limitation of our experimental procedure, suggesting that infants had correctly extracted the label-category pairing during the first trials but the task was too long for them, considering their state of fatigue after running other experiments in our Laboratory.

Finally, we report a linear regression with exposure to a Second Language as a predictor and Change on the PLTT between baseline and post-orientation window as independent variables. This test showed that, at least at 8 months, the percentage of exposure to a second language has no predictive power (R^2= 0.3; P = 0.68).

2.2.3. Discussion

The 8-month-old bilingual and monolingual participants tested in Experiment 1 showed signs of learning the label-category pairing during the familiarization trials. Our procedure does not allow us to discard the possibility that infants activated the reward accidentally, but the fact that they reacted only after listening to the label suggests that they processed the relationship between label and category.

The results of the categorization trials reinforces this conclusion, pointing towards the possibility that infants successfully mapped the label to the category. When we corrected for category preferences, we found indication that a brief familiarization with a label followed by a reward induced 8-month-olds, both bilinguals and monolinguals, to turn towards the target category in the Categorization Trials.

The results also show that the procedure introduced by Saksida (2014) to study the relationship between names and single objects can also induce a label-category pairing in young infants.

A more fine-grained analysis on the temporal course also reinforced this conclusion.

First, when splitting the trials by considering the identity of the target category, we found that infants that had been familiarized with the "non-attractive" category began the trial oriented toward the attractive category. However after listening to the label they were capable of overcoming their initial preferences and oriented towards the target category. When participants were familiarized with the "attractive" category they were already oriented during the Baseline period. This led simply to a prolongation of the same looking pattern.

Then, by analyzing the temporal course of the whole experiment, we found that infants, and especially monolinguals, gave clear signs of learning in the first half of the experiment, that is, in test trials following a very limited amount of familiarization. The fact that the strongest effects were concentrated in the first half of the experiment may suggest two conclusions. The first conclusion is theoretical: at this age the process of labeling can induce a quick response of grouping images into categories. Similar results have been found, but with older infants. For example, Woodward, Markman, and Fitzsimmons (1994) found that 13 and 18-months-old infants rapidly learn the meaning of new names for objects after a brief training. Also, Spiegel and Halberda (2011) showed that toddlers can fast-map novel names even after one single presentation. We believe the current result is the first evidence that much younger infants may also learn such pairings quickly. This leads to the question of what exactly infants learn at 8 months, and of whether the developmental trajectory of label learning is stable or goes through stages that may recruit different mechanisms. The second conclusion is procedural: perhaps the structure of our task was too repetitive to be interesting enough for infants to bear with it until the end of the experiment. It is worth recalling that most infants in our experiment had already participated on one or two other experiments. Thus, it is not impossible that the overall load on infants was excessive and that they could only give due attention at the beginning of the experiment.

Our third conclusion is suggested by the temporal dynamics of the preferences in the test phase. We found that monolinguals' and bilinguals' responses were different. Bilinguals presented a very brief early trend towards learning immediately after the orientation onset, although the trend is too ephemeral to crystalize into a sufficiently long temporal stretch. In contrast, monolinguals reacted more slowly, but more consistently, showing a very strong increase in preferences for the target category at a later moment.

What mechanisms may have driven these differences? One possibility is that the monolingual response was driven by the label and therefore required a longer time to process the information.

Bilinguals, instead, might have responded on the basis of non-linguistic information, perhaps attending to visual features of the stimuli such as shape, and resolving the task independently of the label. However, importantly, the fact that they waited until the label before orienting suggests that they at least minimally processed the label before orienting. Another possibility is that bilinguals were more flexible in disengaging from the first image they were looking at during the baseline period, and were faster in redirecting their attention toward the named image after listening to the label. This possibility is congruent with the bilingual advantage in inhibition found at this age (Kovács & Mehler, 2009a)

We know that bilinguals at this age seem to solve tasks involving an executive function component better, which is needed both in positive categorization as well as in inhibiting a previous predominant response. In Experiment 2, we try to further pursue the importance of executive functions in fast categorization by adding a second component to the categorization task. This component is meant to test how monolinguals and bilinguals can flexibly use a recently acquired category - label pairing to bootstrap the learning to another potential label referent. For this purpose, we used a modified version of a ME task.

2.3. Experiment 2: Categorization and flexibility in label-category learning at 8 months of age

In Experiment 2, we added two fundamental changes to the task introduced in Experiment 1.

First, we tested 8-month-old infants with a shorter version of the Infant Contingent Categorization task. This modification was introduced to try to adapt the paradigm to the results we found in Experiment 1. In that experiment, infants tended to react especially in the first part of the experiment, perhaps suggesting that an adaptation to their state or fatigue or to their interest in the task was needed. At the same time, Experiment 2 will allow us to try and replicate the effects observed in Experiment 1 showing that infants can rapidly map a label to a category.

Second, we added a novel contrast to assess infants' ability to disambiguate the referent of a novel label by using an analog of the mutual exclusivity principle within an incidental learning paradigm. We present a novel label, together with a member of the previous target category and a member of the category not associated with the label heard before. Specifically, we ask whether infants can map the novel label with the category that had not been associated with any label in previous presentations thus showing a remarkable ability to flexibly re-orient their learning strategy. The possibility of finding such a result may be supported by a previous experiment by Saksida (2014). In her studies, with a task similar to the one we implemented borrowing her procedure, as early as at 4 months of age infants could disambiguate the meaning of a new label referring to a new single individual. When presented with a familiar object for which a label had been induced and a novel object, 4-month-olds preferred to look at the novel object.

We will try to extend this and our previous findings by exploring what changes if the experiment establishes an association between categories and labels, rather than between single objects and labels. Possibly, differences among bilinguals and monolinguals may appear. We are interested in exploring differences in the strategy that both groups select to try to solve our analog to the ME contrast, the *Incidental Mutual Exclusivity task.*

If the IME analog taxes infants' flexibility at adjusting an unfolding learning strategy, then bilinguals may succeed and not monolinguals. If instead, the IME task is solved more as a linguistic task, then perhaps bilinguals may accept that a category can have multiple labels while monolinguals may not do it, thus predicting an opposite pattern of results.

Finally, it is also possible that a double label-category association task such as the one we proposed is simply too complex for infants. Thus we may find overall potential success in categorization and failure in IME task, despite the success obtained by Saksida already at 4 months.

2.3.1. Materials and methods

2.3.1.1. Participants

Forty-eight full term 8-month-old infants were retained in the analysis: 26 Monolinguals (14 girls. Mean age: 7;22, Range: 7,05 – 8;26) and 22 Bilinguals (13 girls. Mean age: 7;27, Range 7;01 – 8;26). All participants were healthy and free from birth complications according to the parental reports.

Sixteen infants (6 monolinguals and 10 bilinguals) were tested but excluded from the analysis because they were crying or refused to keep seated (7), they did not meet the inclusion criteria (7) (see § 2.2.2.1.), the infant was sick (1), or the infant moved during the testing session in a way that prevented the Eye tracker from collecting data (1).

The total rejection rate was 32%. We want to highlight that in this experiment most participants were also tested after running one or two tasks previously. None of the previous tasks were related to the topics we assess in this experiment.

The bilingualism definition and the recruiting procedure was the same than in Experiment 1.

Thirteen participants were Catalan monolinguals, 13 were Spanish monolinguals and 22 were Catalan – Spanish bilinguals.

2.3.1.2. Materials and Procedure

Materials and methods were identical to those of Experiment 1 except for the changes described below.

First, we selected a subset of 24 images from Experiment 1: 12 representing fish and 12 representing dinosaurs. The pictures where matched to obtain 8 familiarization pairs and 4 test pairs. Each picture was seen twice across the experiment.

Second, we shortened the length of Experiment 2 to 24 trials. These trials were arranged in 4 blocks containing 4 familiarization trials followed by 2 test trials: a Categorization trial and an IME trial. The total duration of the task was around 2.5 min.

Third, we added to the Familiarization and Categorization trials, identical to those of Experiment 1, novel IME trials.

In *IME Trials,* infants listened to a novel label while they were shown two novel images. One picture belonged to the category that had been associated with a label in the familiarization trials. The second picture belonged to the contrasting category presented in the familiarization trials but had not been associated with any label. As in Categorization trials, target images in IME trials did not loom and had the same timing as Categorization Trials. These trials were always presented after a Categorization Trial, as the order was not alternated. Thus, each block was composed of 6 trials: four familiarization and two test trials (one categorization, one IME)

2.3.2. Results

2.3.2.1. Scoring

Data were obtained and analyzed as in Experiment 1.

Note that in IME trials the Target Category is the opposite as that in the Categorization trials. In the Categorization trials the Target Category is the familiarized category, as it refers to the objects that were paired to a label and could loom in familiarization trials. In the IME trials the Target Category should be the one which has *not* be rewarded during familiarization. Thus, the results in Experiment 2 should be read the following way: if an infant responds to the novel label by looking at the image associated with the label in previous trial presentations, we would expect an increase on the PLTT. By contrast, if an infant looks toward the category that was not paired to any label during the familiarization trials, we will observe a decrease in the PLTT.

The preliminary analysis of the time course in Familiarization trials determined that infants oriented towards the familiarized category at around 1900 ms after the label offset. This value was used to define our Baseline and Post-Label windows of analysis (see § 2.2.2.1.).

2.3.2.2. Results

Preliminary analyses on the Familiarization Trials did not show any effect of sex (Girls or Boys) ($P = 0.74$) or Label (Doti or Mapu) ($P = 0.68$). We did find and effect of Target Category.

A one-way ANOVA on the PLTT in Familiarization Trials with Target Category (Dinosaur/Fish) as a between – participants factor and Participants as a random factor nested in Target Category, revealed a main effect of Target Category ($M\ PLTT_{Dinosaur}$ = 72.14, $M\ PLTT_{Fish}$ = 52.62; $F(1,48) = 68.96$; $P = 0.0001$). Post hoc Scheffé tests revealed infants had a strong preference for the category dinosaur over fish.

As in Experimen 1, to correct for baseline preferences for a particular category, we compared infants' performance, not against chance, but against their baseline preference scores.

2.3.2.2.1. Familiarization Trials

The results reported here include the averaged responses to all the Familiarization trials collapsed.

A two-way mixed ANOVA was calculated on the PLTT in Familiarization trials with Time Window (Baseline/Post-Label) as a within – participants factor, Linguistic background (Bilinguals/ Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background. Results showed a main effect of Time Window ($M\ PLTT_{Baseline}$ = 49.49, $M\ PLTT_{Post\text{-}Label}$ = 74.37; $F(1,46) = 125.31$; $P < 0.0001$). We did not find differences between bilinguals and monolinguals ($P = 0.78$) nor an interaction between Time Window and Linguistic background ($P = 0.29$). Results showed that both bilinguals and monolinguals equally increased their PLTT score during the Post-Label window.

The time-course analysis of the PLTT revealed a clear continuous time sequence in the Post-Label period in which both groups increased their PLTT with respect to their baselines. In monolinguals, the sequence started 770 ms after the label offset and extended for 1800 ms (that is, from 1900 to 3600 ms after the trial onset). In bilinguals, the sequence also started at 770 ms after the label offset and lasted for 1700 ms (that is, from 1900 to 3500 ms after the trial onset).

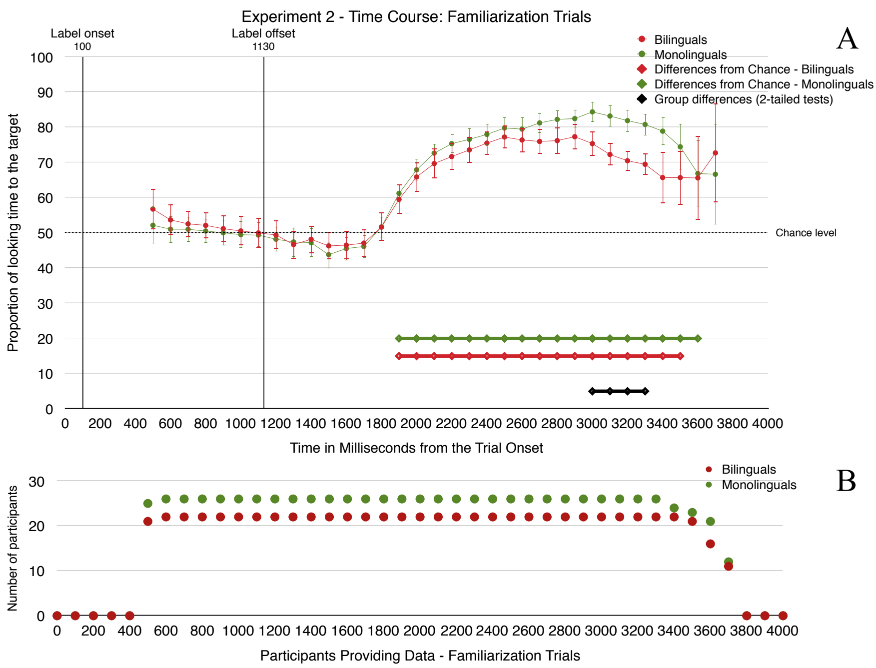


Figure 10. **A.** Symbols representing the PLTT across time in the Familiarization Trials of Experiment 2.. Vertical lines represent the onset and offset of the label. The Horizontal line indicates the Chance level. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from Chance level for the bilingual and monolingual group and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

The analysis did not identify any region in which the time course of bilinguals and monolinguals differed (see Figure 10).

The results of the time-course analysis showed that both groups equally increased their PLTT during the Post-Label period. Results suggest both groups could trigger the reward and that once it started looming it captured infants' attention for the target location.

2.3.2.2.2. Categorization Trials

We first ran an overall analysis, in which we extracted the average responses to the 4 trials of the Categorization condition trials collapsed. For each participant, a Mean PLTT score was calculated for the Baseline and the Post-Label periods. Then, we ran a two-way mixed - factors ANOVA on PLTT in Categorization Trials with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic background (Bilinguals/Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background. Results showed a main effect of Time Window ($M\ PLTT_{Baseline} = 45.76$, $M\ PLTT_{Post\text{-}Label} = 51.76$; $F(1,46) = 4.60$; $P = 0.037$). We did not find an effect of linguistic background ($P = 0.37$) or an interaction between Linguistic Background and Time Window ($P = 0.66$). Results showed that both groups increased their PLTT after the label presentation, suggesting they could identify the target category (see Figure 11).

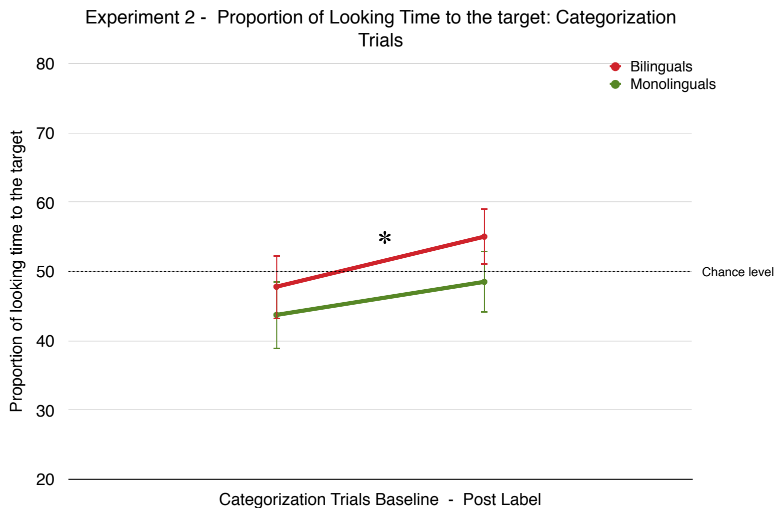


Figure 11. Mean PLTT during the Baseline and Post-Label period of Experiment 2 for Categorization Trials. Bars represent SEs. * indicate effects at $p< 0.05$

Then, we conducted the same analysis with Target Category as an independent variable, because it was shown to have an effect on infants PLTT in previous analyses. We ran a three-way mixed - factors ANOVA on the PLTT in the Categorization Trials with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic Background (Bilinguals/Monolinguals) and Category (Dinosaur/Fish) as a between–participants factor and Participants as a random factor nested in Linguistic background * Category. Results revealed a main effect of Time Window, as seen in the previous analysis and a main effect of Target Category as expected. An interaction between Time Window and Category was found ($F(1,44) = 4.15$; $P = 0.04$). No other interactions were found. The Scheffé test showed that infants who had been familiarized with the fish category presented a significant increase of PLTT in the Post-

Label period (P = 0.004) while the ones that were presented with the dinosaur category as a target did not change their PLTT in the Post-Label-Period (P = 0.93) (see Figure 12).

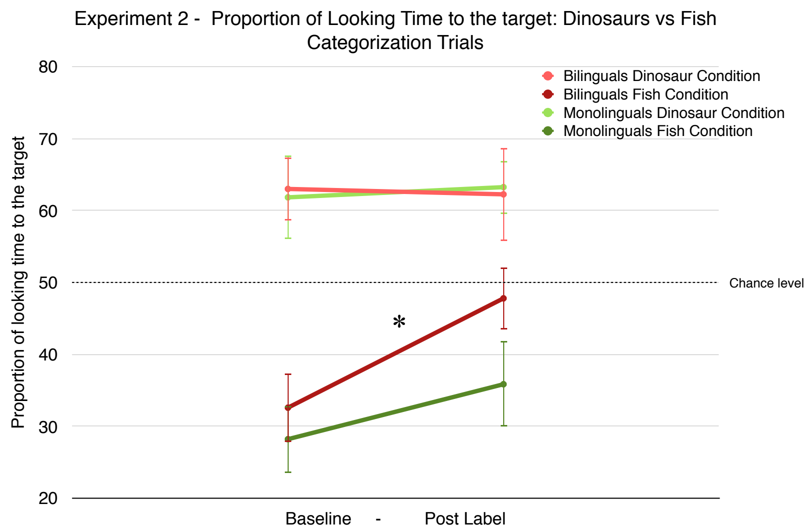


Figure 12. Mean PLTT during the Baseline and Post-Label period of Experiment 2 in Categorization Trials split by Target Category. Bars represent SEs. * indicate effects at $p < 0.05$

Putting both analyses together, we find a pattern results similar to that of Experiment 1. Overall infants had a preference for the dinosaur category. When that category was the target, they maintained their baseline preference in the Post-Label phase. Instead, when the target category was the fish, they overcame this preference and tended to look more at the fish in the Post-Label phase.

Then we assessed the role of presentation repetition during the experiment by analyzing the first and the second half of the experiment separately.

A two-way mixed ANOVA on the PLTT of the first half of Categorization trials with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic Background (Bilinguals/ Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background was conducted. The analysis did not identify a main effect of Time Window (P = 0.32) or Linguistic Background (P = 0.33) nor an interaction (P = 0.51).

However, the same ANOVA on the second half of Categorization Trials revealed a main effect of Time Window (M $PLTT_{Baseline}$ = 45.50, M $PLTT_{Post\text{-}Label}$ = 55.39; $F(1,46)$= 6.19; $P = 0.017$), showing that infants increased their looking at the target category after the orientation point. There were no a main effects of Linguistic Background (P = 0.34) nor an interaction (P = 0.86) (see Figure 13).

Figure 13. Mean PLTT during the Baseline and Post-Label period for first and second halves of Categorization Trials of Experiment 2. Bars represent SEs. * indicate effects at p< 0.05

This result suggests that when the learning task is more taxing, learning effects still appear, but require a firmer (although always short) exposure to the relationship between label and reward for infants to grasp it.

We then ran a Time-course analysis. Each time-bin in a Post-Label window of 1 s was compared against the Mean PLTT score for the Baseline period. An independent baseline score for the monolingual and bilingual groups was obtained.

The temporal course analysis of the PLTT in Categorization trials identified an early time window in monolinguals and a late time window in bilinguals in which infants increased their PLTT after the label. For monolinguals, this period extended from 200 ms to 400 ms after the orientation onset (300 ms). For the bilingual group, it started 800 ms after the orientation onset and ended 1200 ms after the orientation onset (500 ms). There were no differences between monolinguals and bilinguals in any region of the temporal course of the trials (see Figure 14).

Because the time window of the effects for bilinguals falls partially outside of the predefined temporal window of interest for the analyses, we decided to conduct the same time-course analysis extending the temporal window of interest from 1 s to 1.5 s. Following Guthrie and Buchwald (1991) procedure, the results for a sample of N=20, at a significance level of $P < 0.05$, assuming the highest autocorrelation and a window of 15 time points were still significant.

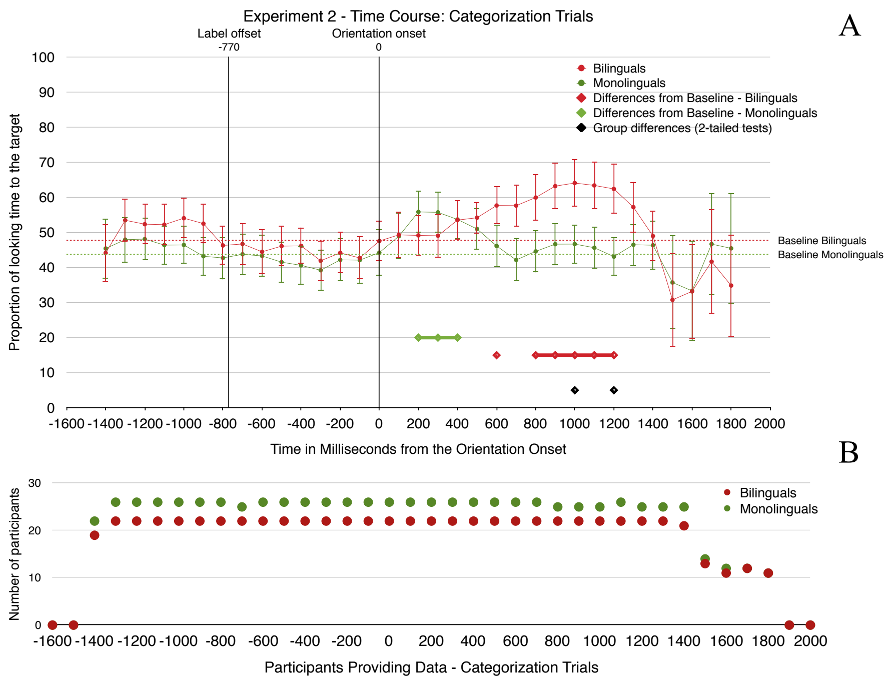


Figure 14. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 2. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

The time-course analysis split by Target Category showed that infants who were familiarized to the fish category, started the trial looking at the non-target category and increased their PLTT after the orientation onset. The analysis did no show any window in which monolinguals increased the PLTT. For bilinguals it was found a sequence from 400 ms to 700 ms after the orientation onset (400 ms) in which they increased the PLTT. No group differences were

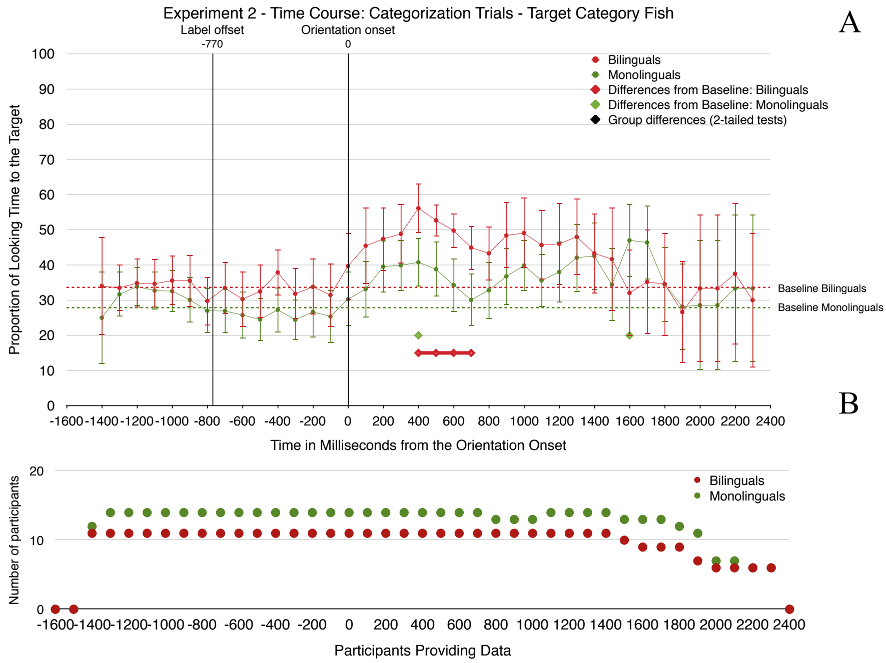


Figure15. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 2 for infants familiarized with the Target Category Fish. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p< 0.05$ different from baseline in monolinguals and bilinguals and the significant sequence of two-tailed t-test at $p< 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

found (see Figure 15).

The time-course analysis of the infants who were familiarized with the category dinosaurs, revealed they started the trial oriented towards the target category. The mean PLTT during the Baseline period was above chance in the monolingual (Mean PLTT = 64.10; t

(11) = 2.46; P = 0.03) and bilingual group (Mean PLTT = 66.81; t (10) = 3.38; P = 0.007).

Despite the high initial baseline scores, a marginal effect was found in bilinguals. Bilinguals increased their PLTT around 1000 ms after the orientation onset. It was a brief sequence (300 ms) that do not reach our significance criteria The effect appeared at a later time window than for infants familiarized with the Target Category Fish (see Figure 16).

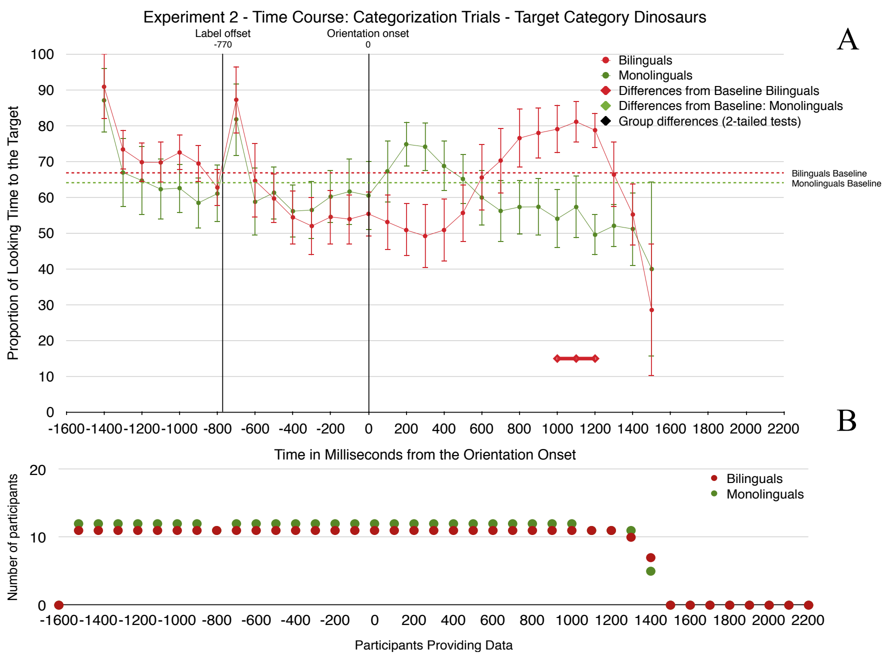


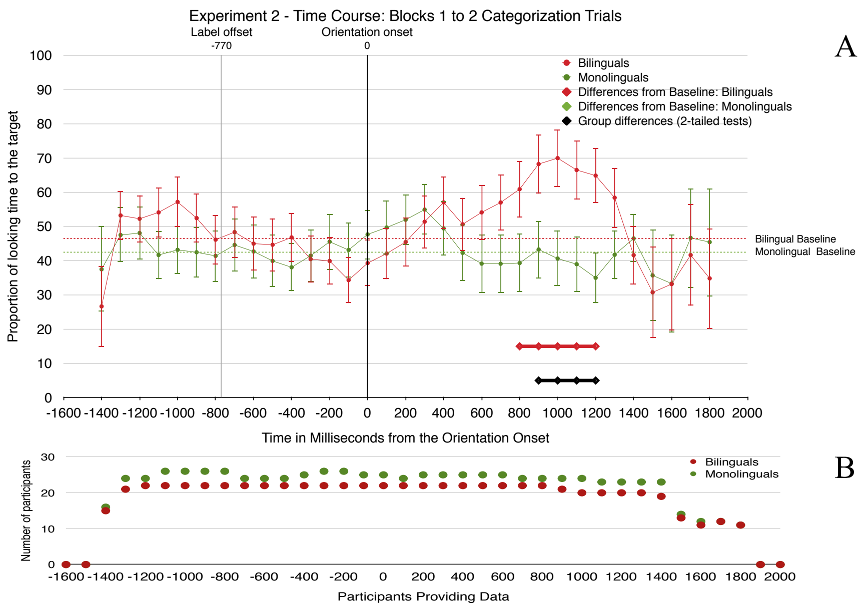
Figure16. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 2 for infants familiarized with the Target Category Dinosaurs. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

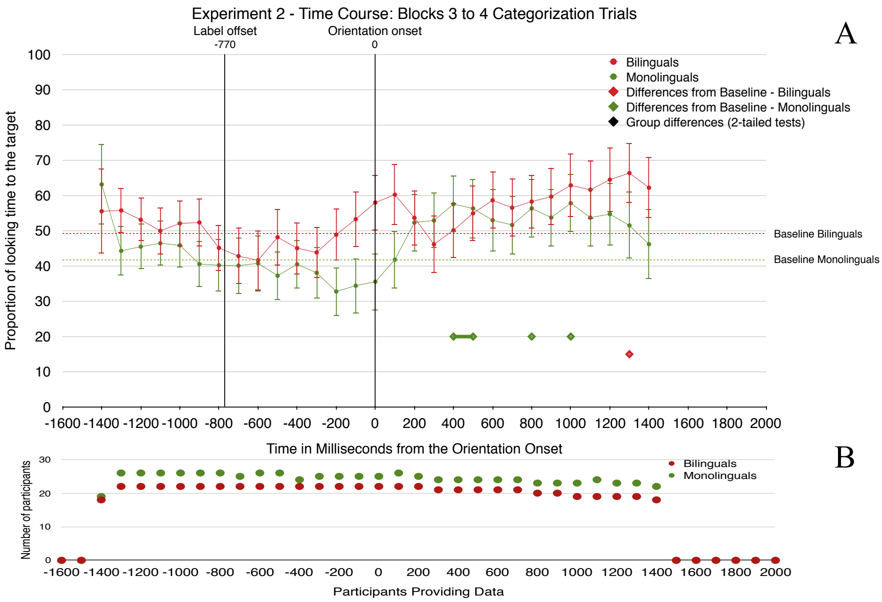
The results split by Target Category suggest that bilingual infants overcame their initial preferences and reacted to the label increasing their PLTT if they were not oriented to the target image (Fish category) or they maintained the attention on the Target Category (Dinosaur category) if they where already oriented.

To assess the progression of learning during the experiment, we ran of the first half (Blocks 1-2) and the second half of experiment (Blocks 3-4). The bilingual group showed learning signs already in the first half of the task. Bilinguals increased PLTT with respect to the baseline between 800 ms until about 1200 ms after the orientation onset. Monolinguals did not. Indeed, in that interval bilinguals and monolinguals differed significantly separate analyses (see Figure 17).

For monolinguals, learning did not appear until the second half of the experiment although not significantly according to our criterion (see Figure 18).

Finally, a linear regression with Second Language as a predictor and Change in the PLTT between baseline and post-orientation window as an independent variable found that, as in Experiment 1, the percentage of exposure to a second language has no predictive power ($R^2 = 0.1$; $P = 0.86$).





Figures 17-18. **A.** Symbols representing the PLTT across time in the first and second half of Categorization Trials of Experiment 2. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

2.3.2.2.2. Incidental Mutual Exclusivity Trials

We first ran an overall analysis, in which we extracted the average response to the 4 trials of the IME condition collapsed.

We ran a two-way mixed-factors ANOVA on the PLTT in IME Trials with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic background (Bilinguals/ Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background. We did not find a main effect of Time Window (P = 0.54) or Linguistic Background (P = 0.57) or an interaction (P = 0.52). Infants did not modify their PLTT after listening to the label in IME trials (see Figure 19).

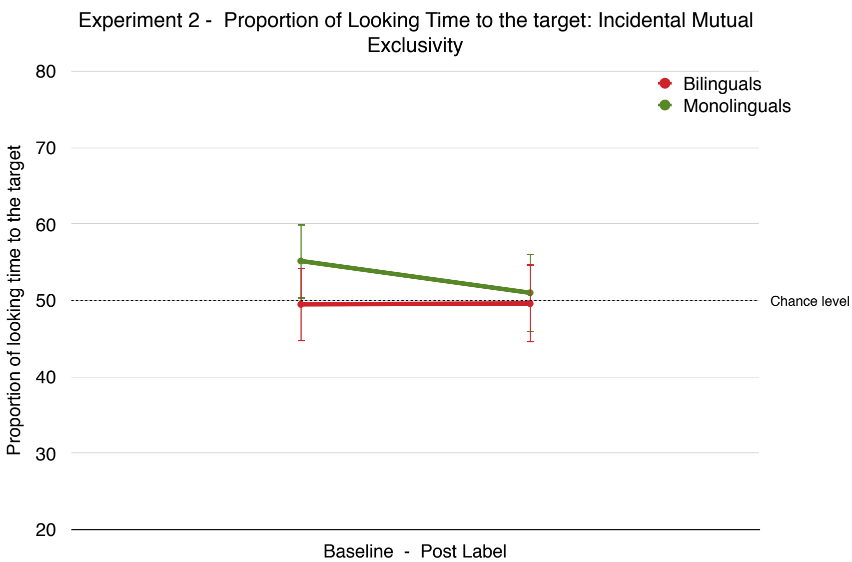


Figure 19. Mean PLTT during the Baseline and Post-Label period of the IME Trials of Experiment 2. Bars represent SEs. * indicate effects at p< 0.05

Additionally, we ran the same ANOVA including the variable Target Category since it was shown to have an effect on the PLTT in previous analysis. A three-way mixed ANOVA was calculated on the PLTT in the IME Trials with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic Background (Bilinguals/Monolinguals) and Category (Dinosaur/Fish) as a between – participants factor and Participants as a random factor nested in Linguistic background * Category was calculated. Results did not identify any main effect, except for Category as seen in the preliminary analysis ($P = 0.0001$). There was no interaction between Time Window and Category ($P = 0.97$).

An analysis of the first and second half of the IME trials separately to asses the progression of PLTT across presentation repetitions did not reveal any significant effects either.

We then performed a Time-course analysis of PLTT in IME trials. The temporal analysis did not identify any temporal interval in which any group differed from their baselines or from each other. (see Figure 20).

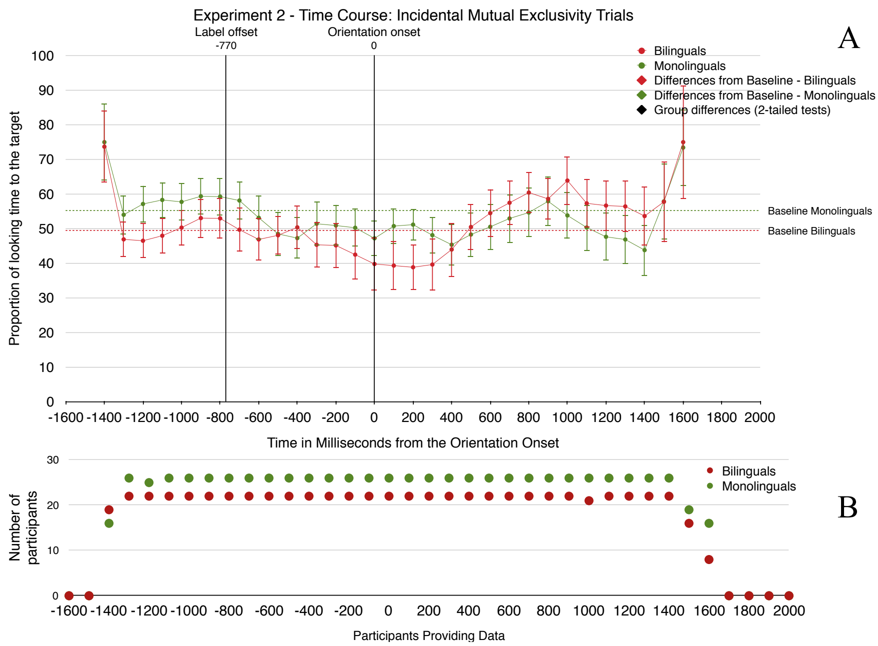


Figure 20. **A.** Symbols representing the PLTT across time in IME Trials of Experiment 2. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Separate time-course analyses were run for each Target Category type. The analyses identified an early window in which bilingual infants familiarized with the fish category reduced their PLTT after listening to the label in IME trials. The sequence started at the orientation onset and lasted for 400 ms (see Figure 21)

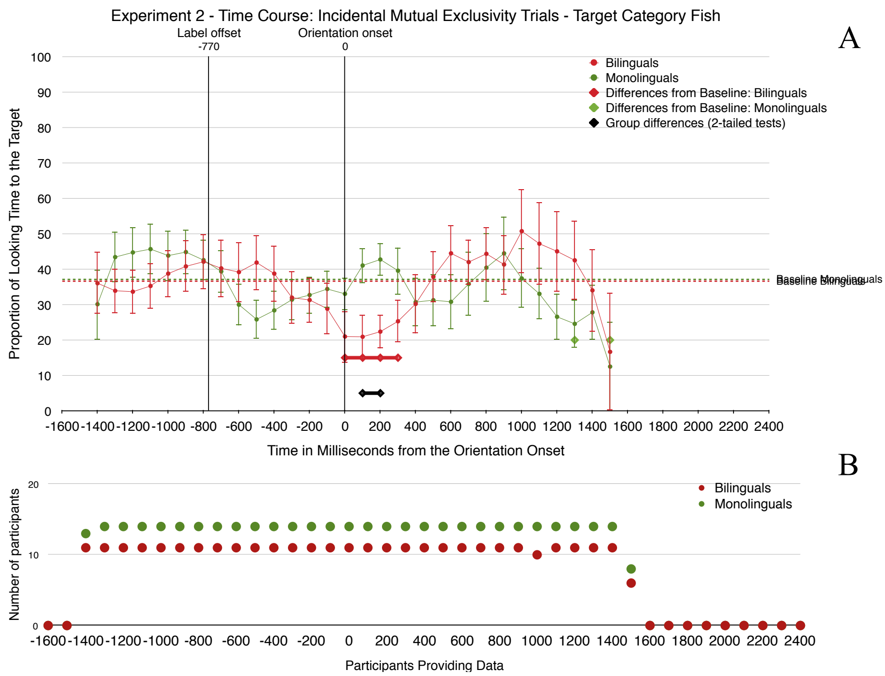


Figure 21. **A.** Symbols representing the PLTT across time in the IME Trials of Experiment 2 for the infants familiarized with the Target Category Fish. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p< 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p< 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Results did not find any window in which bilinguals or monolinguals familiarized with the category dinosaurs decreased the PLTT after the label or differed from each other (see Figure 22)

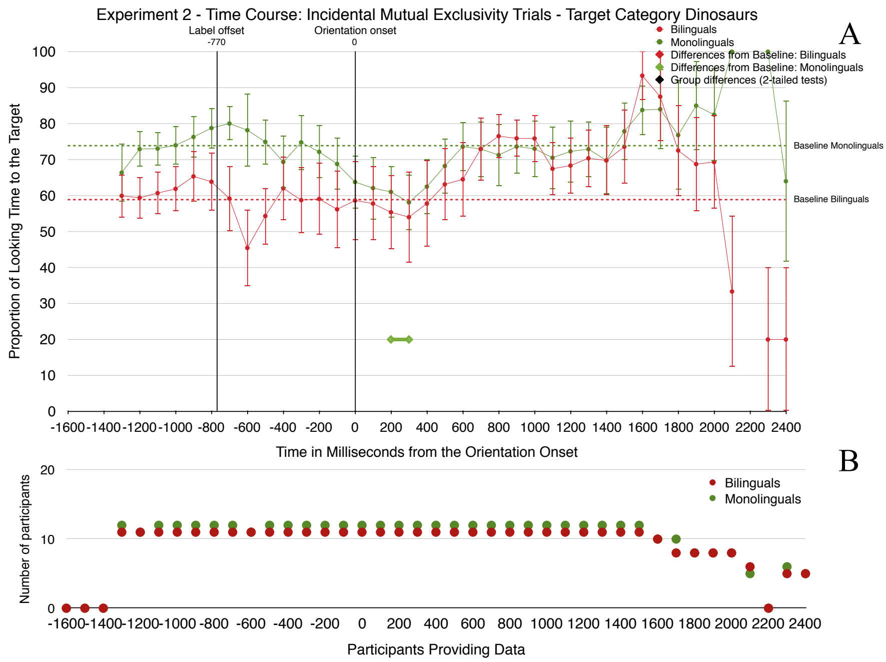


Figure 22. **A.** Symbols representing the PLTT across time in the IME Trials of Experiment 2 for the infants familiarized with the Target Category Dinosaurs. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Interestingly, when we analyzed the temporal course of the first half of the IME trials (Blocks 1-2) separately from the second half of test trials (Blocks 3-4), monolinguals oriented towards the target non-target category (our measure of success) during the the second half of task. This occurred between the orientation onset and 500 ms after it. Nothing more than a very slight tendency appeared in bilinguals (not significant according to our criteria). No differences emerged between bilinguals and monolinguals suggesting that the behavior of both groups was not strikingly different (see figure 23).

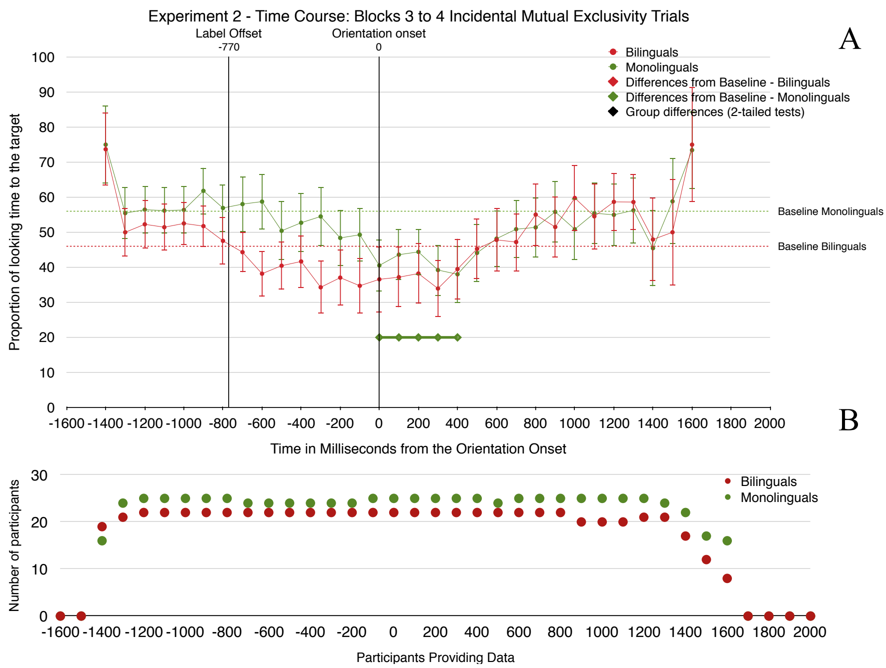


Figure 23. **A**. Symbols representing the PLTT across time in the second half of the IME Trials of Experiment 2. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Finally, a simple linear regression was calculated to predict the Change on the PLTT based on the Percentage of Exposure to a

Second Language. The variable Percentage of Exposure had no predictive value ($R^2 = 0.1$; $P = 0.87$).

2.3.3. Discussion

The results of Experiment 2 showed that both groups of 8-months old infants significantly increased their PLTT during the Post-Label period of familiarization trials. This pattern of results parallels that of Experiment 1 and suggests that both groups could successfully trigger the reward and that once it was active infants tended to keep looking at the target location. This result is important because it shows that, despite inserting two non-reinforced test trials within the familiarization trials (compared to only one non-reinforced trial in Experiment 1), infants still from triggered the reward.

The time-course analysis of the familiarization trials identified a late window in which the dynamics of responses differed between bilinguals and monolinguals. The fact that bilingual infants presented lower scores compared to monolinguals at the end of the trials might suggest that bilingual infants lost interest in the task faster than the monolingual group.

In the categorization trials, the ANOVA showed a main effect of time window indicating that infants in both groups increased their proportion of looking time towards the target after the label. This result suggests that infants responded to the label by increasing their PLTT and, therefore, could have successfully extracted the contingency between the label and the target Category. This result replicates the findings in Experiment 1.

However, when we computed separate ANOVAS on the PLTT for the first and second half of the Categorization trials, we found that it was not until the second half of the experiment that infants significantly increased their PLTT. In Experiment 1, we had found that infants already started to increase their PLTT in Categorization trials during the first half of the experiment. However, we need to point out that because in Experiment 1 there were more trial repetitions, at the end of the first half of experiment 1, infants had been presented with more trials than at the end of the first half of Experiment 2. Still, it is possible that by adding a second type of test contrast in Experiment 2 we increased the task complexity. It might be possible that infants in Experiment 2 needed more evidence before starting to respond.

Surprisingly, the analysis of the time course of Categorization trials revealed a categorization advantage for bilinguals that we had not seen when we analyzed the averaged responses of the Baseline and the Post-Label period. The time-course analysis showed that bilinguals in the first half of the experiment and just after a brief training could fast map a label to a visual category. Monolinguals, instead, did not start to give signs of Categorization until the second half of the Experiment. This result contrasts with the monolingual Categorization abilities shown in Experiment 1.

In the IME trials, we did not find significant results when we compared the average Baseline and Post-Label scores. Again, when we looked at the data with a more fine-grained analysis, such as the time-course analysis, significant effects emerged. Although we did not find changes in the PLTT during the first half of the IME trials, we found that by the end of the second half, monolinguals significantly decreased their PLTT after listening to a novel label. Results might suggest that monolinguals mapped the novel label with the visual category that had not received a name during the familiarization. The fact that we only observe this effect at the end of the experiment has different possible interpretations. One possibility is that the IME contrast is more demanding, thus needing more trials before being able to succeed. Another more interesting possibility is that the mastering of the IME contrast is constructed on the bases of some previous knowledge. Monolingual infants may have needed to first understand the target category-familiar label mapping. Then, once they have identified the referent of the first label, they can map the novel label to the remaining reference candidate. If that was the case, the question remains of why monolinguals infants did not show behavioral sign of comprehending the target category-familiar label pairing.

Moreover, we saw that bilinguals rapidly extracted the label - category pairing in Categorization trials but did not show clear signs of solving the IME contrast. Therefore, extracting the first category-label pairing cannot be the only factor in the resolution of the IME trials. We speculate this pattern of results might indicate that bilinguals fast extracted the category-label contingency in Categorization trials thanks to their executive advantage. However, in a more linguistically driven task as the IME, bilinguals who are used to listening more than one name for the same object, do not have sufficient evidence to disambiguate the referent of the new label and did not show a unitary response. By contrast, monolinguals more used to listening a single name per object, perhaps rejected the target category and preferred to attend to the category member that had not received a name yet.

In experiment 3, we will try to disambiguate the role of attentional and linguistic factors by testing 15-months-old bilinguals and monolinguals with the same task. At that age, infants might have accumulated more experience with their languages and iIf IME requires a certain degree of linguistic knowledge, we might now see both groups showing a preference in the IME trials.

2.4. Experiment 3: Categorization and flexibility in label-category learning at 15 months of age

In Experiment 3, we presented 15-month-old monolingual and bilingual infants with the Infant Contingent Categorization and Incidental Mutual Exclusivity tasks. The aim of this experiment is to explore if older infants, who have higher cognitive resources and a deeper knowledge of their linguistic environment, may perform better. But there are other reasons why testing infants at this age is particularly interesting. When an 8-month-old infant is faced with tasks such as our Categorization or IME task, it is not clear how she reconstructs the relationship between a sound and its associated label. It is doubtful that, at an age at which infants do not even make the difference between a sound of their language and a sound of a foreign language, they can construct a sound label as a full lexical However, at around 12 months a restructuring of the relationship between sound and phonemes occurs, at least for monolinguals. After that period, infants could, in principle, treat sounds as full words, at least because they have a phonological representation available for building the representation of a lexical entry. Thus, perhaps a switch between treating sounds as labels, and treating them as words, may occur between 12 and 15 months. Interestingly, at 15 months, monolingual infants have already converged onto the phonological repertoire of their language, but bilingual infants have not entry (Pi, 2015). Hence, monolinguals can interpret the Categorization task as a real word learning tasks, but perhaps bilinguals cannot. Thus, language difference can also arise for our IME task. Moreover, controversial findings have been reported about the ability to use ME at that age. There is some evidence suggesting that 15-months-old monolinguals can use the ME constrain (Markman et al., 2003) but there is also evidence suggesting that monolinguals do not use the ME principle until older ages (Halberda, 2003).

2.4.1. Materials and methods

Materials and methods of Experiment 3 were identical to those of Experiment 2.

2.4.1.1. Participants

Fifty full term, 15-month-old infants were retained for analysis: 30 Monolinguals (18 girls. Mean age: 15;09, Range: 14;20 – 15;27) and 20 Bilinguals (11 girls. Mean age: 15;03, Range 14;17 – 16;00). All participants were healthy and free from birth complications according to parental report.

Thirteen infants (5 monolinguals and 7 bilinguals) were tested but excluded from the analysis because they cried or refused to keep seated (6); they did not meet the inclusion criteria (5) (see § 2.2.2.1.); they had a trilingual linguistic profile (1); or the calibration failed (1).

The total rejection rate was 21%. In this experiment, some participants were tested after previously running other experiments not related to the current tasks.

The bilingualism definition and the recruiting procedure were equal to those of Experiment 1. Sixteen participants were Catalan monolinguals, 14 were Spanish monolinguals and 20 were Catalan – Spanish bilinguals.

2.4.1.2. Materials and Procedure

The materials and procedure used in Experiment 3 were identical to those of Experiment 2.

2.4.2. Results

2.4.2.1. Scoring

Data were obtained and analyzed as in Experiment 2.

The preliminary time course analysis of the Familiarization trials showed that infants oriented towards the target category for the first time around 670 ms after the label presentation. That value was used to define our Baseline and Post-Label windows of analysis (see § 2.2.2.1.).

2.4.2.2. Results

The preliminary analysis on the Familiarization trials did not reveal effects of sex (Girls or Boys) ($P = 0.71$) or label (Doti or Mapu) ($P = 0.37$). But we did find an effect of Target Category.

A one-way ANOVA was run on the PLTT in Familiarization Trials with Target Category (Dinosaur/Fish) as a between – participants factor and Participants as a random factor nested in Target Category. This analysis, revealed a main effect of Target Category (*M PLTT*

Dinosaur = 75.25, *M PLTT* *Fish* = 47.17; F(1,50) = 94.10; P = 0.0001). Post hoc Scheffé tests revealed infants had a strong preference for the category Dinosaur over Fish. As in previous experiments, to correct for baseline preferences for one category, we compared infant's performance, not against chance, but against their baseline preference scores.

2.4.2.2.1. Familiarization Trials

The results reported here include the collapsed responses to all Familiarization trials.

For each participant, a Mean PLTT score was obtained for the Baseline and the Post-Label period. A two-way mixed-effects ANOVA was run on the PLTT in Familiarization Trials with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic background (Bilinguals/Monolinguals) as a between–participants factor and Participants as a random factor nested in Linguistic background. Results showed a main effect of Time Window (*M PLTT* *Baseline* = 51.83, *M PLTT* *Post-Label* = 71.71; F (1,48) = 59.32; P = 0.0001). There were no effects of Linguistic Background (P = 0.99) nor an interaction between Linguistic Background and Time Window (P = 0.58). The results show that both groups equally increased their PLTT during the period following the label.

We also conducted a time-course analysis to localize the dynamics of the effects. The analysis identified two time windows in which the two groups increased the PLTT above chance. For monolinguals, it started 670 ms after the label offset and lasted around 1800 ms (that is, from 1800 ms to 3500 ms after the trial onset). For bilinguals, the sequence also started 670 ms after the label offset and lasted for 2100 ms (that is, from 1800 ms to 3800 ms after the trial onset). There were no differences between the temporal course of the two groups (see Figure 24).

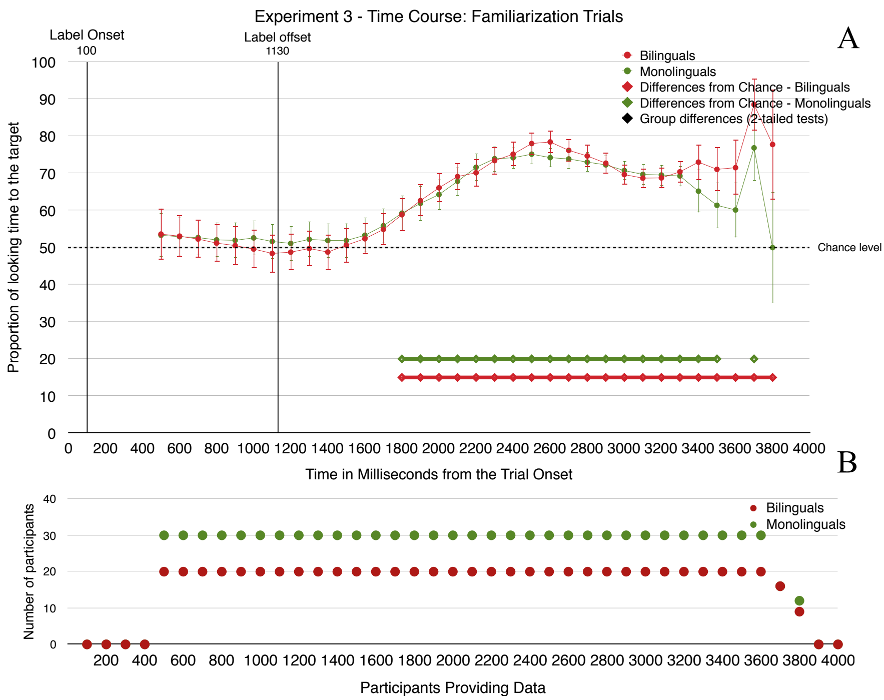


Figure 24. Symbols representing the PLTT across time in the Familiarization Trials of Experiment 3. Vertical lines represent the onset and offset of the label. The Horizontal line indicates the Chance level. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from Chance level for the bilingual and monolingual group and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

The results of the Familiarization trials showed that bilinguals and monolinguals significantly increased their PLTT during the Post-Label period. It suggests that both groups could trigger the reward and that they tended to kept looking at it once the movement started.

2.4.2.2.2. Categorization Trials

We first ran an overall analysis, in which we extracted the average of the responses to the 4 trials of the Categorization condition collapsed. For each participant, Mean PLTT scores were calculated for each time windows of interest (Baseline and Post-Label). We then ran a two-way mixed-effect ANOVA on the PLTT in Categorization Trials with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic background (Bilinguals/ Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background. Results did not show a main effect of time window ($P = 0.82$) or linguistic background ($P = 0.91$), or an interaction between Linguistic Background and Time Window ($P = 0.40$). Overall, PLTT did not increase following the label presentation (see figure 25).

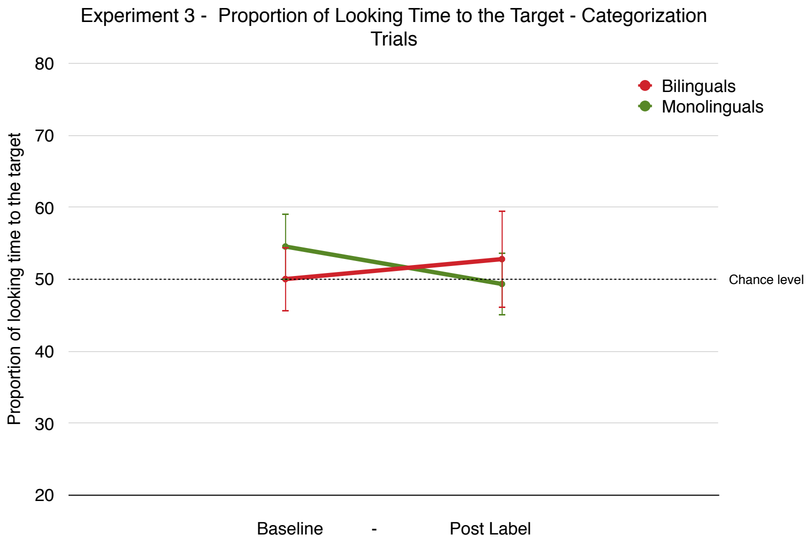


Figure 25. Mean PLTT during the Baseline and Post-Label period of the Categorization trials of Experiment 3. Bars represent SEs. * indicate effects at $p<0.05$

We also conducted the same analysis including Target Category as an independent variable because it was shown to have an effect on infants' PLTT in previous analyses. A three-way mixed-effect ANOVA was run on the PLTT in the Categorization Trials with Time Window (Baseline/Post-Label) as a within – participants factor, Linguistic Background (Bilinguals/Monolinguals) and Category (Dinosaur/Fish) as a between–participants factor and Participants as a random factor nested in Linguistic background * Category. Results did not identify any main effects except for Target Category, revealing an overall preference for the dinosaur category. A marginal interaction between Time Window and Category was found (P = 0.07). No other interaction was significant (see Figure 26).

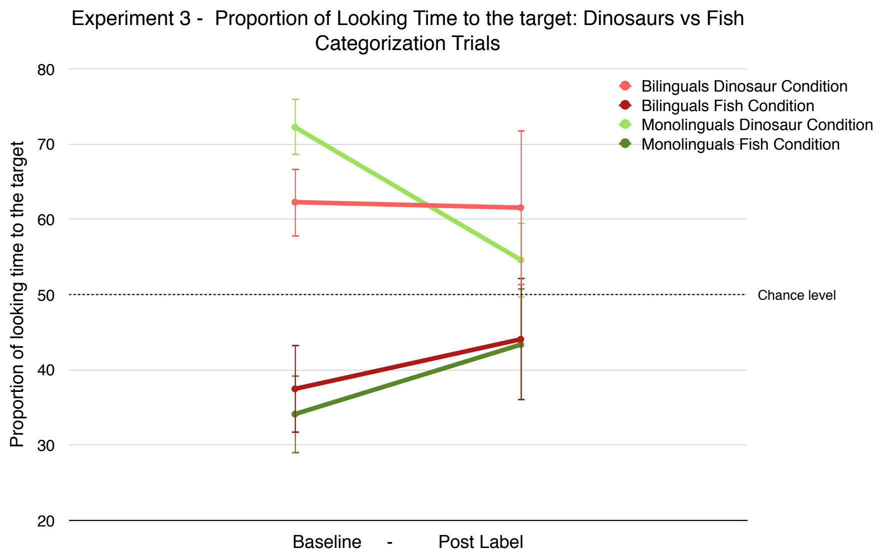


Figure 26. Mean PLTT during the Baseline and Post-Label period of the Categorization Trials of Experiment 3. Bars represent SEs. * indicate effects at $p < 0.05$

Two separate ANOVAS were conducted on the first and second halves of the Categorization trials in order to assess the progression of learning across trials. No main effects or interactions were found.

Then, to explore possible changes on PLTT across the temporal course of the task we ran a time-course analysis of the Categorization Trials. Each time-bin in the Post-Label window of 1 s was compared against the mean PLTT score for the Baseline period. The temporal analysis did not identify any interval in which any group differed from their baselines or from each other. The results showed that neither group increased the PLTT during the Post-Label period (see Figure 27).

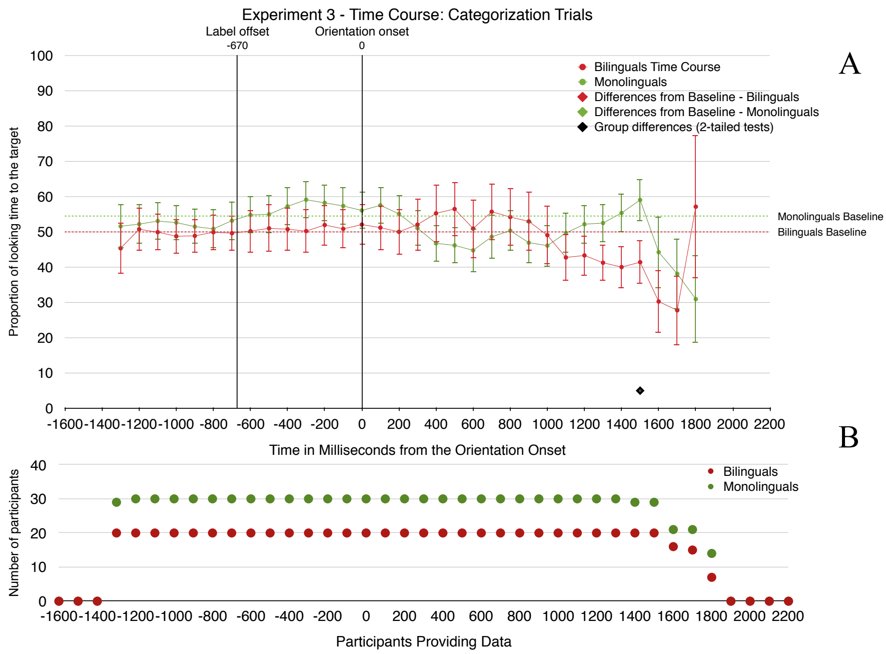


Figure 27. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 3. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p< 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p< 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Since previous analyses discovered an effect of Target Category on infants' PLTT, we performed separate time-course analyses splitting the infants who had been familiarized with the Category Fish from those who had been familiarized with the Target Category Dinosaurs. Monolingual infants who saw the Target Category Fish reinforced in familiarization trials, presented long stretch (800 ms) in which they increased the PLTT scores above the baseline reference starting 1 s after the orientation onset. Bilinguals did not increased their PLTT. No group differences were found (see Figure 28).

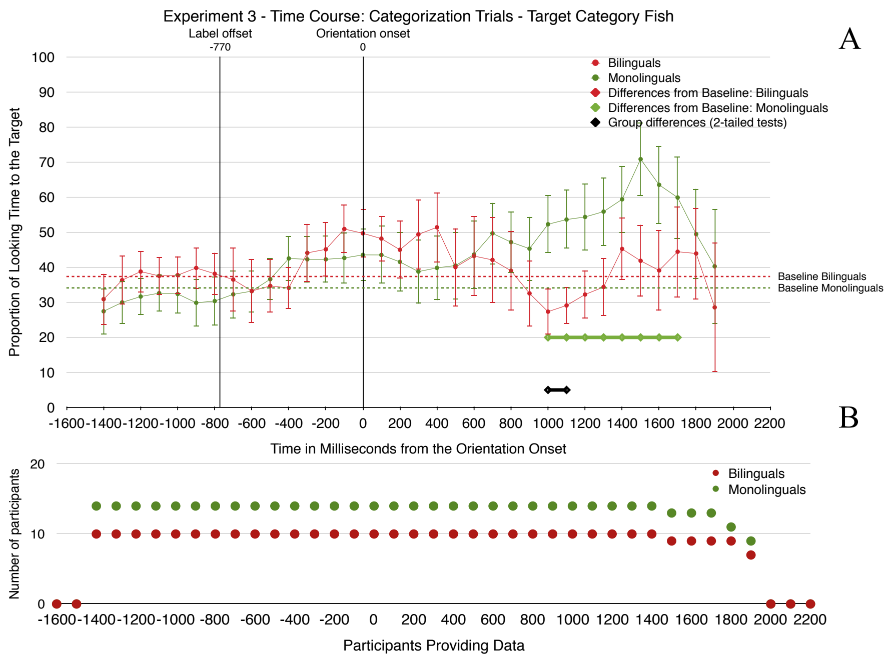


Figure 28. **A.** Symbols representing the PLTT across time in the Categorization Trials of Experiment 3 for infants familiarized with the Target Category fish. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

The time-course analysis of the infants that had been familiarized with the target category Dinosaur, showed that infants started the trial already oriented to the dinosaur category. The mean PLTT during the Baseline period was above chance in the monolingual (Mean PLTT = 72.74; t (15) = 6.02; P = 0.0001) and bilingual group (Mean PLTT = 63.23; t (9) = 3.03; P = 0.01). In the Post-Label no significant strings were found indicating an increase of PLTT in bilinguals or monolinguals (see Figure 29).

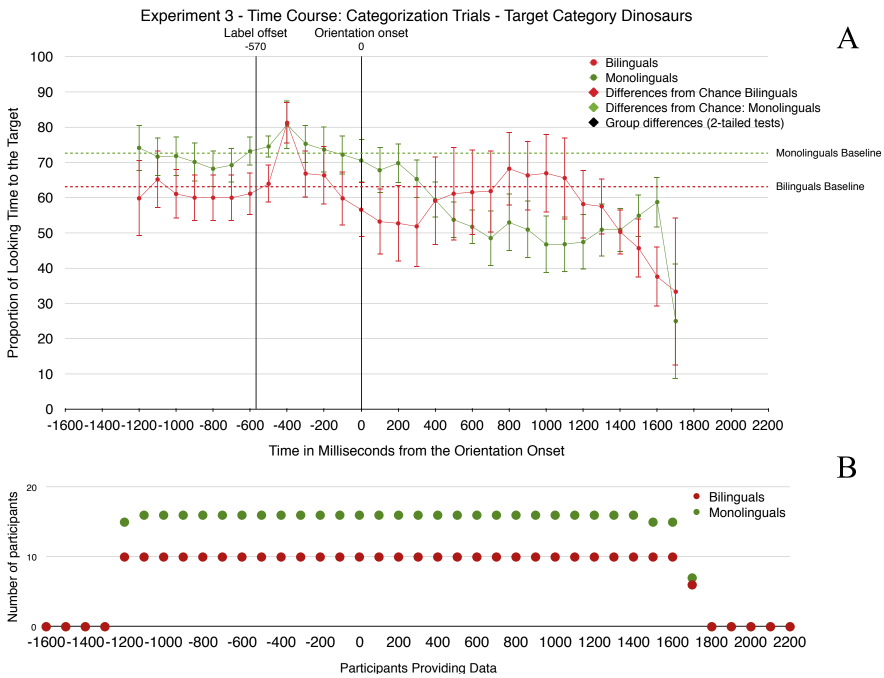


Figure 29. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 3 for infants familiarized with the Target Category Dinosaurs. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Overall, the results of the temporal course split by Target Category revealed a differential pattern for the two categories. Only monolinguals who saw the Target Category Fish in familiarization trials, increased the PLTT during the Post-Label period. Both, bilinguals and monolinguals, that saw the Target Category Dinosaur in familiarization trials started the trials already oriented towards the target picture and no significant increase of the PLTT was found in the Post-Label window, probably due to the initial high scores.

We then conducted a temporal analysis splitting the first (Blocks 1-2) and the second halves (Blocks 3-4) of the Categorization trials to asses the progression of learning across the experiment. The analysis did not identify any significant effects.

Finally, we ran a linear regression with the Percentage of exposure to a second language as a predictor and the Change on the PLTT between Post-Label and Baseline period as the dependent variable. Results showed that at 15-moths of age, the Percentage of exposure to a second language did not predict performance in our Categorization task ($R^2 = 0.9$; $P = 0.50$).

2.4.2.2.3. Incidental Mutual Exclusivity Trials

We first ran an overall analysis, in which we extracted the average response in the IME condition collapsed—i.e. collapsing 4 trials. A two-way mixed-effects ANOVA was conducted on PLTT in IME Trials with Time Window (Baseline/Post-Label) as a within – participants factor, Linguistic Background (Bilinguals/

Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background. Results did not show a main effect of Time Window (P = 0.41) or Linguistic Background (P = 0.47). There was no interaction between Linguistic Background and Time Window (P = 0.31). Results showed that bilinguals and monolinguals did not modify their PLTT during the Post-Label period (see Figure 30).

Then, we conducted the same analysis including the variable Target

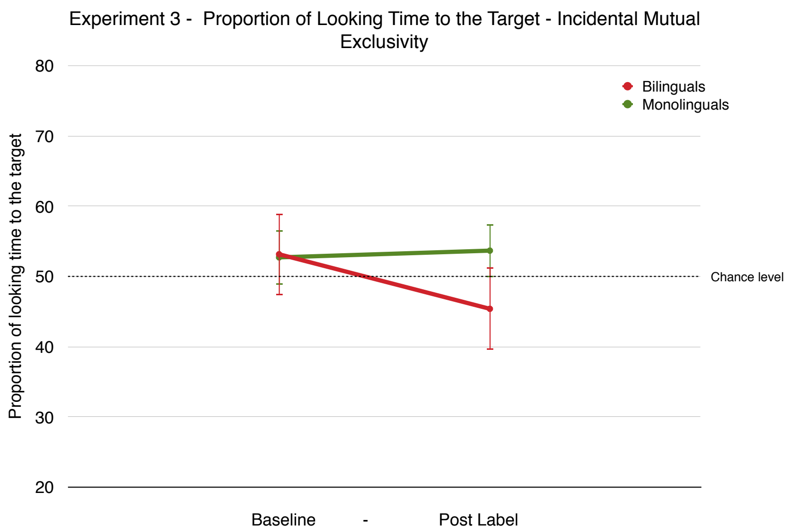


Figure 30. Mean PLTT during the Baseline and Post-Label period of the IME Trials of Experiment 3. Bars represent SEs. * indicate effects at $p < 0.05$

Category, which was shown to have an effect in the PLTT. We ran a three-way mixed ANOVA on the PLTT in the IME Trials with Time Window (Baseline/Post-Label) as a within – participants factor, Linguistic Background (Bilinguals/Monolinguals) and Category (Dinosaur/Fish) as a between – participants factor and Participants as a random factor nested in Linguistic background * Category. Results did not show any main effects, except for Category, as expected. An interaction between Time Window and Category emerged ($F(1,46) = 7.52$; $P = 0.009$). Post hoc Scheffé tests revealed that infants who had been familiarized with the target category Dinosaur tended to reduce their PLTT during the Post-Label period ($P = 0.01$) while infants that had been familiarized with the target category fish did not change their PLTT scores ($P = 0.19$). Overall, the two analyses suggest that infants did not modify their PLTT after listening to the label (see Figure 31)

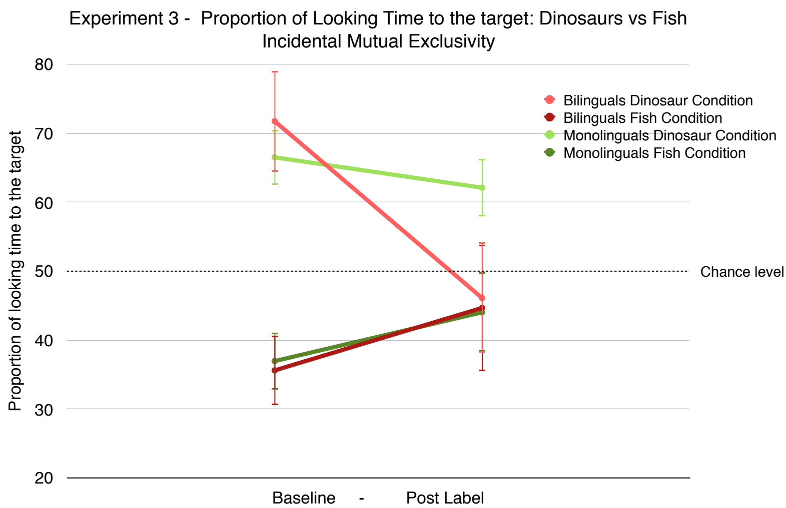


Figure 31. Mean PLTT during the Baseline and Post-Label period in the IME Trials of Experiment 3 split by Target Category. Bars represent SEs. * indicate effects at $p < 0.05$

To study the progression of learning through repetitions, we ran two separate analyses on the the first half and second halves of IME trials. The analyses did not reveal any main effects or interaction.

Then, we computed a time-course analysis on the PLTT in the IME Trials. The analysis did not identify any sufficiently long sequence of t-test in monolinguals or bilinguals. There were no significant group differences either. Results showed that in IME trials, infants did not reduce their PLTT after listening to a novel label (see Figure 32).

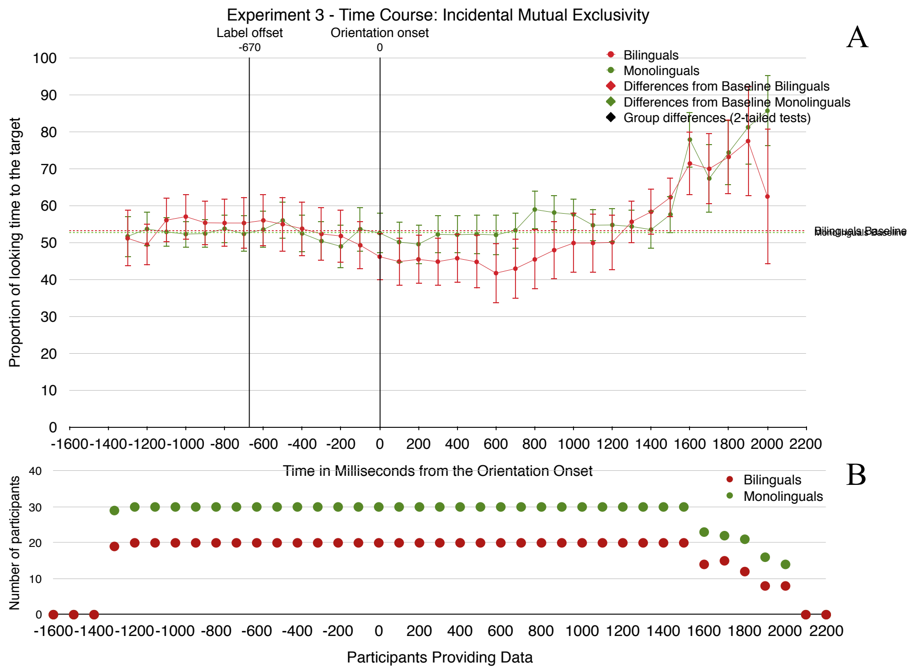


Figure 32. **A.** Symbols representing the PLTT across time in the IME Trials of Experiment 3. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Separate time-course analyses were calculated for the each category type, because target category was shown to have an effect on the PLTT. Infants who were familiarized with the target category Fish, overall looked at the stimuli for less time overall than those who saw the category Dinosaur. Results did not reveal any time window in which bilinguals or monolinguals familiarized with the target category Fish changed their PLTT after the label. No group differences were found (see Figure 33).

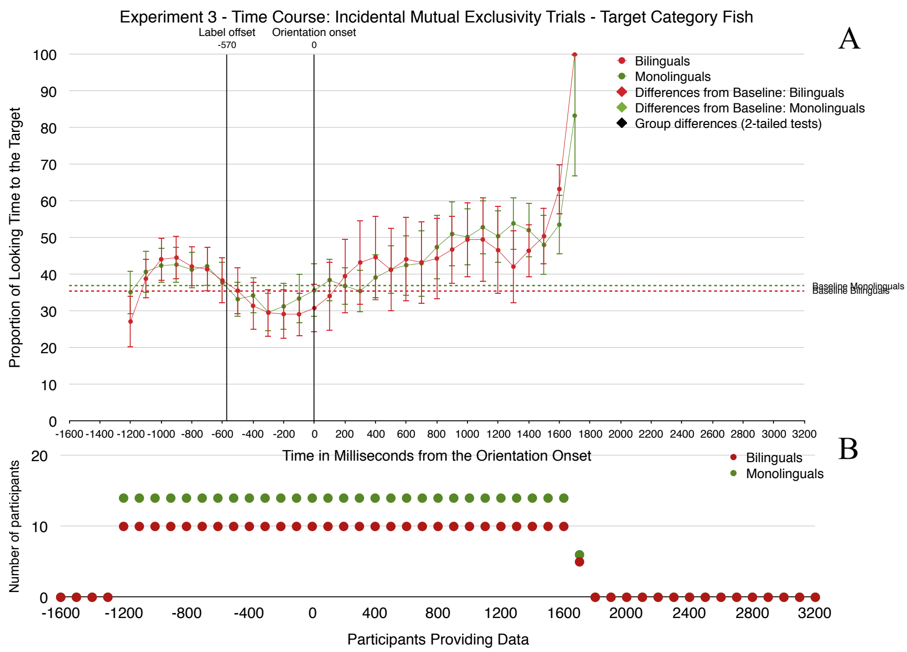


Figure 33. Symbols representing the PLTT across time in the IME Trials of Experiment 3 for infants familiarized with the Target Category Fish. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

The time-course analysis of infants familiarized with the target category Dinosaur showed that they started the trial already oriented to the dinosaur category. The mean PLTT during the Baseline period was above chance in the monolingual (Mean PLTT = 68.48; $t(15) = 4.67$; $P = 0.0003$) and bilingual groups (Mean PLTT = 73.81; $t(9) = 3.3$; $P = 0.009$).

In the Post-Label period, an early time window appeared in which bilinguals clearly reduced their PLTT. This sequence started 100 ms after the orientation onset and lasted for around 1200 ms. In monolinguals, a brief time window emerged (300 ms) in which infants reduced the PLTT. It was a late effect, starting around 1200 ms after the orientation onset. There were no group differences (see Figure 34).

Then, we computed separate time-course analyses for the first (Blocks 1-2) and second halves (Blocks 3-4) of IME trials. In the first half, no reduction appeared on the PLTT of either group, nor were there group differences.

In the analysis of the second half, it was found a short stretch (300 ms) in which monolinguals infants reduced their PLTT. That was not significant according to our criteria, The sequence started 100 ms after the orientation onset and lasted 300 ms. No other window or group differences were identified

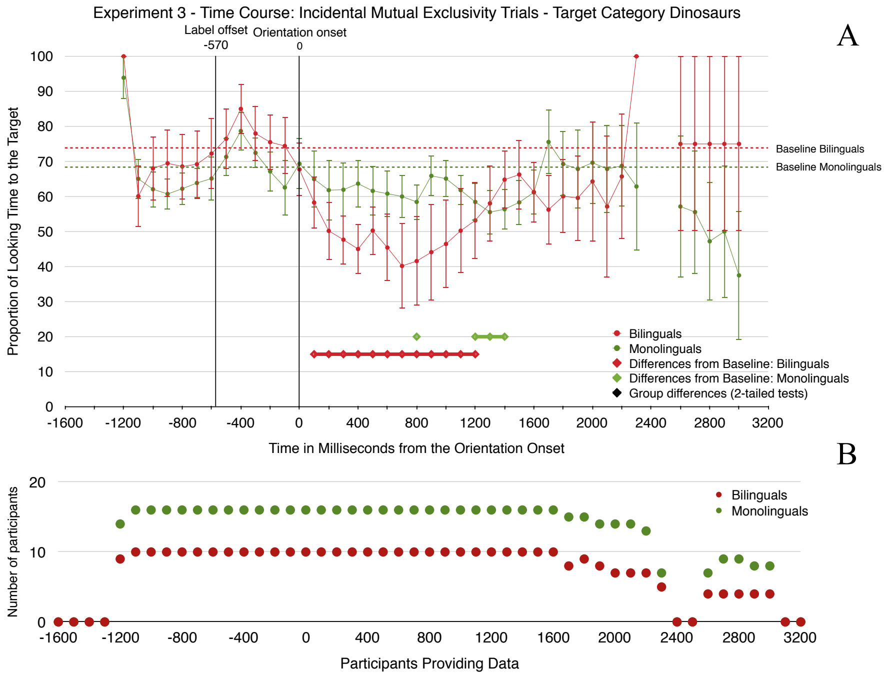


Figure 34. **A.** Symbols representing the PLTT across time in the IME Trials of Experiment 3 for infants familiarized with the Target Category Dinosaur. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Finally, a linear regression with Percentage of Exposure to a Second Language as a Predictor and Change in PLTT between the Post-Label and the Baseline period as a dependent variable was calculated. The regression found that, at 15-month of age, the Percentage of Exposure to a Second Language did not predict infants Change in PLTT in IME trials ($R^2 = 1.3$; $P = 0.42$).

2.4.3. Discussion

In Experiment 3, both 15-month-old bilingual and monolingual infants presented an increase on the PLTT during the Post-Label window of the familiarization trials. The increase was identical in bilinguals and monolinguals. Results suggested that both groups equally triggered the reward and that once the looming was active infants tended to keep looking at the target location.

In Categorization Trials, neither bilinguals nor monolinguals showed clear signs of learning the label category-pairing.

In IME trials, we did not find signs of mapping the novel label to the category that had not been labelled, as neither bilinguals nor monolinguals reduced the PLTT in the Post-Label period. Yet, monolinguals showed a trend during the last trials of the experiment that may suggest they perhaps solved the task independently of the learning of the label-category pairing in Categorization trials.

Overall, the results of this experiment, do not show any systematic patterns of responses at 15-months of age that clearly indicate that infants learned the label-category association in Categorization trials or that they could use this new learning to boost the mapping of a second label association in the IME trials. Results did not allow us to extend the ME results found in Saksida (2014) for single object-label pairs to categories. Surprisingly, 15-month-olds were less responsive to our task than were 8-month-old infants. Perhaps this fact is due to methodological aspects of the tasks. It might be that the kind of fast presentation of the label-category pairing with few repetitions of this association captured 8-month-olds' more superficial, but more immediate reactions, but not those of 15-month-olds. It is possible that the different degree of lexical development that certainly distinguishes 8-month-old from 15-months-old infants may render the task more difficult at 15 than at 8 months. That is, it is possible that 8-months-old infants see the task as a simple label-pair association task, without treating labels as real lexical entries. By contrast, 15-months-old infants may interpret the task as a real word-learning situation, thus asking for more evidence before committing to the relationship between a word and its meaning.

Another possibility is that 8-months-old infants respond to local relationships between sounds and categories—i.e. to short sequences of sound-picture contingencies as independent from each other, whereas 15-months-old participants might try to extract the relationship between the word and its referent by tracking the global cross-situational probability (Smith & Yu, 2008). Indeed, our task, shows at every block four familiarization trials which are reinforced and two test trials which are not. It presents a situation in which, locally, the statistical relationship between labels and pictures is strong, but globally, it is interrupted by the absence of reinforcement for two trials out of every 6. Perhaps 15-months-olds resist associating labels (or words) to category pictures when the label-picture relationship is so frequently interrupted.

In Experiment 4, we will test a group of 19-month-old bilingual and monolingual infants. The behavior of older infants may elucidate the failure found at 15 months. If expertise with words and word-object relations influences the performance in our tasks, we may find that 19-month-old, infants may respond better to them, having accumulated enough linguistic knowledge as to improve their performance. Under this hypothesis, it is also possible that differences in language experience, such as those which characterize monolingual and bilingual infants, may reappear. If, instead, the failure at 15 months is mostly due to methodological aspects of our tasks which may favor younger infants over the older ones, then testing infants at an older age, regardless of their linguistic background, will be ineffective.

2.5. Experiment 4: Categorization and flexibility in label-category learning at 19 months of age

In this experiment, we tested 19-month-old infants with the Infant Contingent Categorization and IME Tasks. There is agreement that around this age, infants tend to notably increase their linguistic abilities. Their comprehension improves. Especially, their vocabulary size sharply increases, preparing the phenomenon that is known as a Vocabulary Burst or explosion (Benedict, 1979).

Other aspects of infants' linguistic development make testing 19-month-olds informative. Around this age, there is evidence that infants can apply the ME strategy for word learning (Halberda, 2003; Houston−Price et al., 2010). The present study is aimed to explore whether more linguistic experience will help 19-month-olds old to solve the Infant Contingent Categorization Task and the Incidental Mutual Exclusivity Task. If that was the case, we may reveal a sort of U-shaped curve between 8 and 19 months in how incidental evidence can be used in word acquisition. At the same time, performance in the IME task may reveal a developmental shift in the ability to disambiguate the referent of a novel label by exploiting recently acquired label-category associations. Potential differences between bilinguals and monolinguals may also indicate whether, at this age, a distinct linguistic experience may shape the optimal word-learning strategy according to the kind of input those infants are exposed to.

2.5.1. Materials and methods

Materials and methods of Experiment 4 were identical to those of Experiments 2 and 3.

2.5.1.1. Participants

Thirty-seven full term 19-months-old infants were retained for analysis: 18 Monolinguals (6 girls. Mean age: 19;02, Range: 18;07 – 20;01) and 20 Bilinguals (13 girls. Mean age: 19:08, Range 18:10 – 20;06). All participants were healthy and free from birth complications according to parental report.

Seven infants (7 monolinguals) were tested but excluded from the analyses because they were crying or refused to keep seated (5); they did not meet the inclusion criteria (1) (see § 2.2.2.1.); or the language questionnaire revealed that they had not listened to neither Catalan nor Spanish (1).

The total rejection rate was 16%. In this experiment, only a few participants were tested after participating to other experimental tasks unrelated with the topics we assessed in this experiment.

Eleven participants were Catalan monolinguals, 7 were Spanish monolinguals, 18 were Catalan – Spanish bilinguals, and 1 was a Spanish - German bilingual.

2.5.1.3. Materials and Procedure

Materials and methods of Experiment 4 were identical to those of Experiments 2 and 3.

2.5.2. Results

2.5.2.1. Scoring

Data were obtained and analyzed as in Experiments 2 and 3.

The preliminary analysis of the time course in Familiarization trials determined that infants oriented towards the familiarized category at around 770 ms after the label offset. This value was used to define our Baseline and Post-Label windows of analysis (see § 2.2.2.1.).

2.5.2.2. Results

Preliminary analyses on the Familiarization trials did not detected any effect of Sex (Girls or Boys) ($P = 0.95$) or Label (Doti or Mapu) ($P = 0.93$). But we did find an effect of Target Category.

A one-way ANOVA on the PLTT in Familiarization Trials with Target Category (Dinosaur/Fish) as a between – participants factor and Participants as a random factor nested in Target Category, revealed a main effect of Target Category ($M\ PLTT_{Dinosaur} = 74.83$, $M\ PLTT_{Fish} = 43.85$; $F(1,37) = 73.29$; $P < 0.0001$). Post hoc Scheffé tests revealed infants had a strong for the dinosaur category over fish. As in previous experiments, we compared infant's performance against their baseline period scores, to correct for baseline preferences for one category.

2.5.2.2.1. Familiarization Trials

We ran an overall analysis, in which we extracted the average of the response to the Familiarization trials collapsed.

A two-way mixed-effect ANOVA was run on the PLTT in Familiarization Trials with Time Window (Baseline/Post-Label) as a within – participants factor, Linguistic background (Bilinguals/ Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background. This test revealed a significant effect of Time Window. Results showed that infants increased their PLTT during the Post-Label window ($M\ PLTT_{Baseline} = 50.92$, $M\ PLTT_{Post\text{-}Label} = 67.09$; $F\ (1,35) = 59.32$; $P < 0.0001$). There was no effect of Linguistic Background ($P = 0.35$) or an interaction between Linguistic Background and Time Window ($P = 0.25$). These results show that both groups equally increased their PLTT during the Post-Label period of the familiarization trials.

Additionally, we conducted a time-course analysis. Both bilinguals and monolinguals increased the PLTT at about the same times. In monolinguals, the sequence started 670 ms after the label offset (from 1800 ms to 3300 ms after the label offset). Bilinguals showed the first signs of orientation slightly later, around 970 ms after the label offset and stayed fixated on the target a bit longer than monolinguals (from 2100 to 3900 ms after the label offset). The analysis did not identify any region in which the time course of bilinguals and monolinguals differed significantly (see Figure 35).

The results of the two analyses agreed that bilinguals and monolinguals increased their PLTT during the Post-Label period. Results suggest both groups could trigger the reward and that once it started looming it captured infants' attention to the target location.

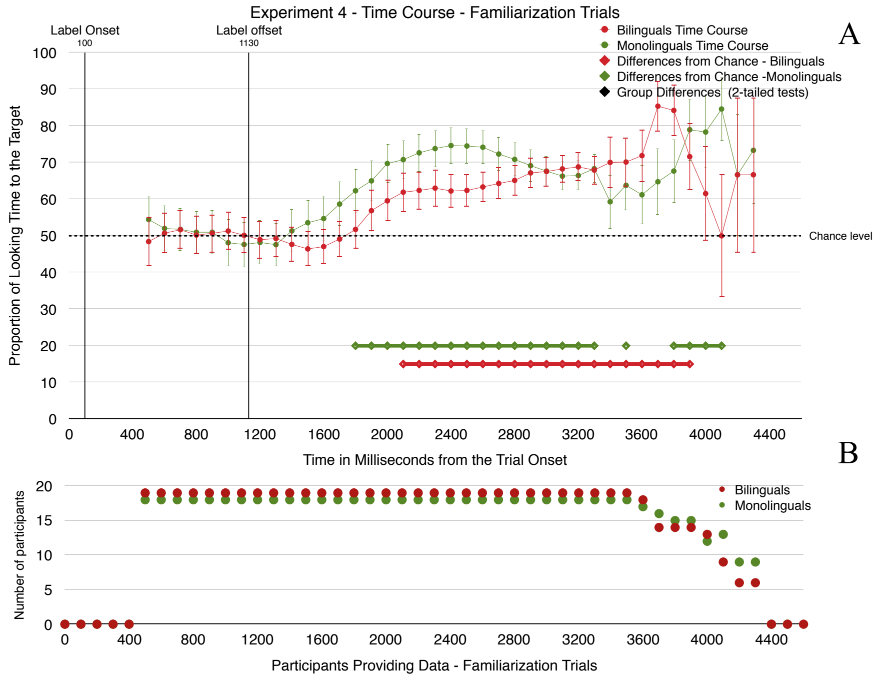


Figure 35. Symbols representing the PLTT across time in the Familiarization Trials of Experiment 4. Vertical lines represent the onset and offset of the label. The Horizontal line indicates the Chance level. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from Chance level for the bilingual and monolingual group and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

2.5.2.2.2. Categorization Trials

We first ran an overall analysis of the 4 Categorization trials collapsed. For each participant a Mean PLTT score was calculated for the Baseline and Post-Label period. Then we ran a two-way mixed ANOVA on the PLTT in Categorization Trials with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic background (Bilinguals/Monolinguals) as a between–

participants factor and Participants as a random factor nested in Linguistic background. Results did not detected any effects of Time Window (P = 0.63) or Linguistic Background (P = 0.69) on the PLTT. However, there was a marginal interaction of Time Window and Linguistic Background (P = 0.07) (see Figure 36).

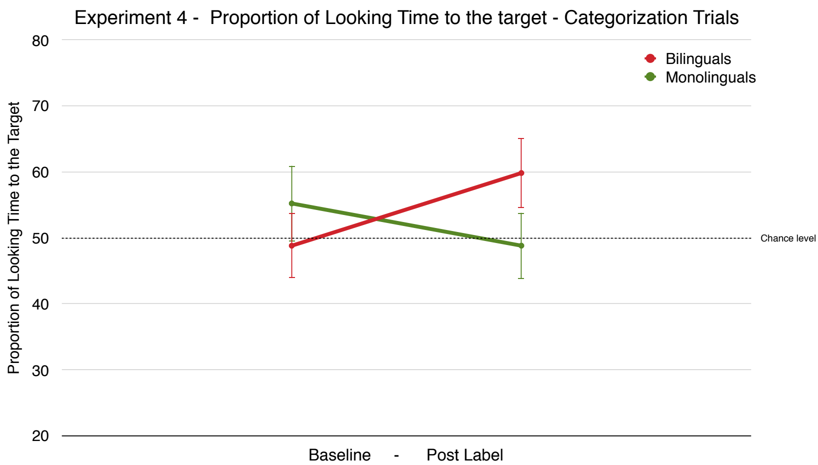


Figure 36. Mean PLTT during the Baseline and Post-Label period of the Categorization Trials of Experiment 4. Bars represent SEs. * indicate effects at p<0.05

The same analysis but including the variable Target Category was run, since this was shown to have an effect on infants' PLTT in the previous analyses.

A three-way mixed-effect ANOVA was run on the PLTT in the Categorization Trials with Time Window (Baseline/Post-Label) as a within – participants factor, Linguistic Background (Bilinguals/ Monolinguals) and Category (Dinosaur/Fish) as a between– participants factor and Participants as a random factor nested in

Linguistic background * Category. Results revealed main effects of Time Window and a main effect of Target Category, as expected. An interaction between Time Window and Category was found (F(1,33) = 9.10; P = 0.004), The Scheffé tests showed that infants who had been familiarized with the fish category presented a significant increase of PLTT in the Post-Label period (P = 0.02) while the ones that were presented with the dinosaur category as a target had a trend towards reducing the PLTT during the Post-Label period (P = 0.08). There was also a marginal interaction between Time Window and Linguistic Background (F(1,33) = 3.74; P = 0.06). No other interactions were found (see Figure 37).

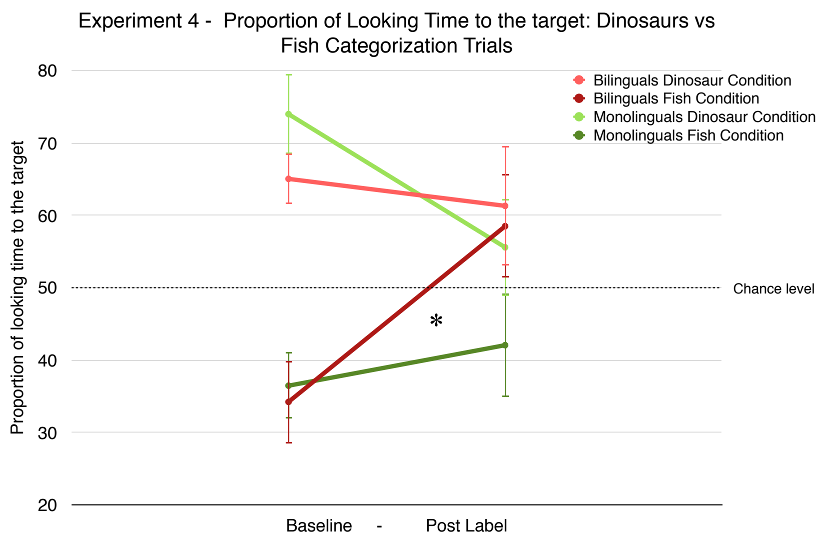


Figure 37. Mean PLTT during the Baseline and Post-Label period of the Categorization Trials of Experiment 4 split by Target Category. Bars represent SEs. * indicate effects at $p < 0.05$

As in previous experiments, we found a preference for the dinosaur category. When dinosaurs was also the Target Category, infants tended to maintain or slightly decrease the PLTT during the Post-Label period. When the target category was fish, infants exhibited the opposite pattern and increased their PLTT after the label.

To study the progression of learning across the experiment, we ran restricted analyses on the first (Blocks 1-2) and second halves (Blocks 3- 4) of the categorization trials. Results uncovered that the increase on the PLTT emerged in the last test trials only. A two-way mixed-effect ANOVA was run on the PLTT in Categorization Trials of Blocks 3 and 4 with Time Window (Baseline/Post-Label) as a within–participants factor, Linguistic background (Bilinguals/ Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background. Results showed a marginal interaction between Time Window and Linguistic Background (P = 0.08). We did not find any other significant effects (see Figure 38).

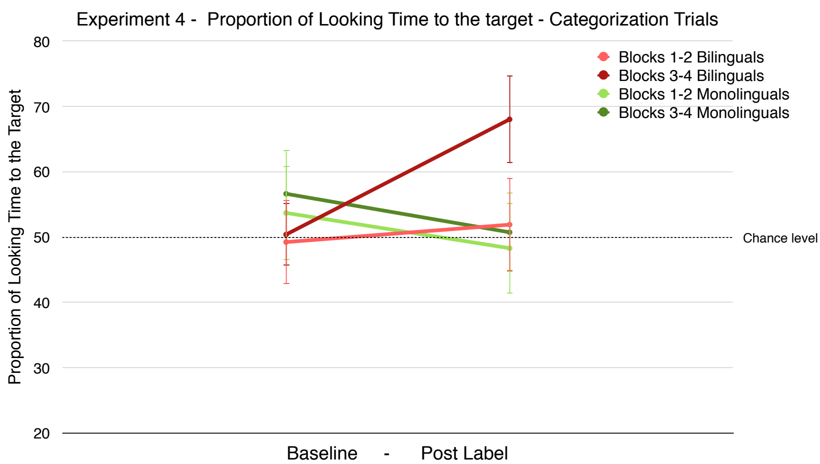


Figure 38. Mean PLTT during the Baseline and Post-Label period split by first and second half of the Categorization Trials of Experiment 4. Bars represent SEs. * indicate effects at $p < 0.05$

Results suggest that for the learning situation was demanding for infants, and that the extraction of the label-category association required a notable amount of brief exposures.

Additionally, a time-course analysis was conducted to better characterize the dynamics of the effects. Each time bin in the Post-Label window of one second was compared against the Mean PLTT Baseline score of each group. The bilingual group exhibited a period of time in which PLTT increased in the Post-Label period (500 ms). The time window of the effect went from 500 ms to 900 ms after the orientation onset. No such effect appeared for the monolingual group, nor differences between the dynamics of the responses of bilinguals and monolinguals (see Figure 39).

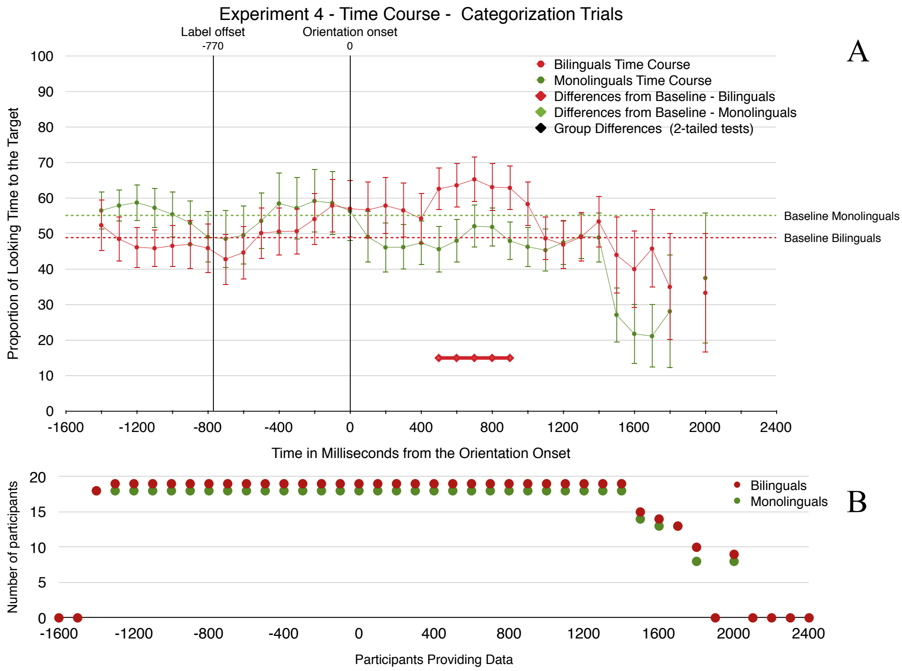


Figure 39. Symbols representing the PLTT across time in the Categorization Trials of Experiment 4. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Because previous analyses had discovered an effect of Target Category on infants' PLTT, we additionally calculated separate time-course analyses, splitting participants according to the reinforced category seen in the familiarization trials.

When infants had been familiarized with the target category fish, bilinguals and monolinguals showed an increase of the PLTT after the label presentation. We also found that infants maintained their attention to the target and non-target ports longer than with Dinosaurs as the target category. For monolinguals, the time period was short (300 ms), starting 800 ms after the orientation onset. Bilinguals presented a clearer increase of the PLTT starting 300 ms and extending to 1000 ms after the orientation onset. No significant group differences where observed (see Figure 40).

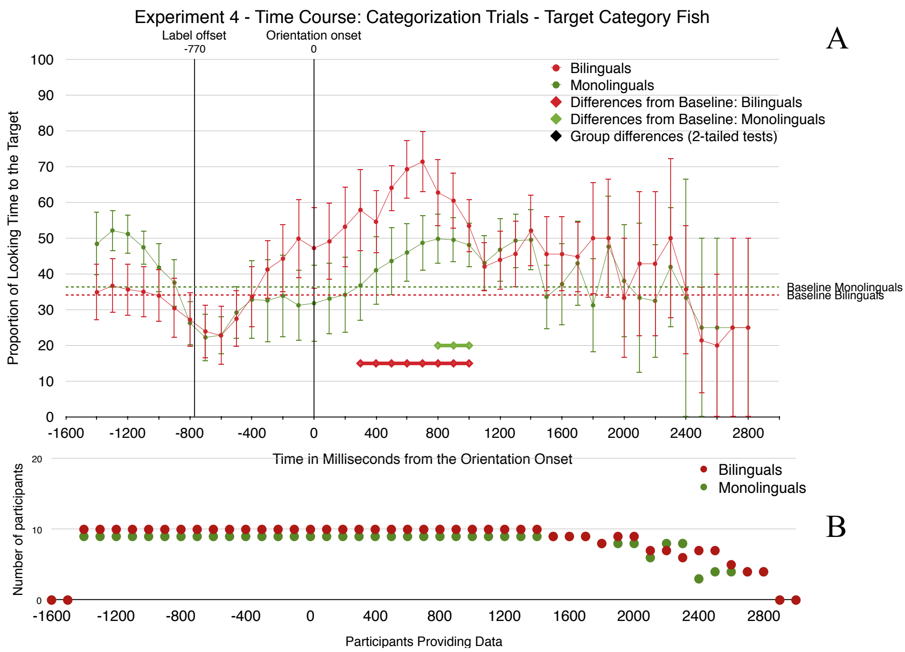


Figure 40. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 4 for infants familiarized with the Target Category Fish. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

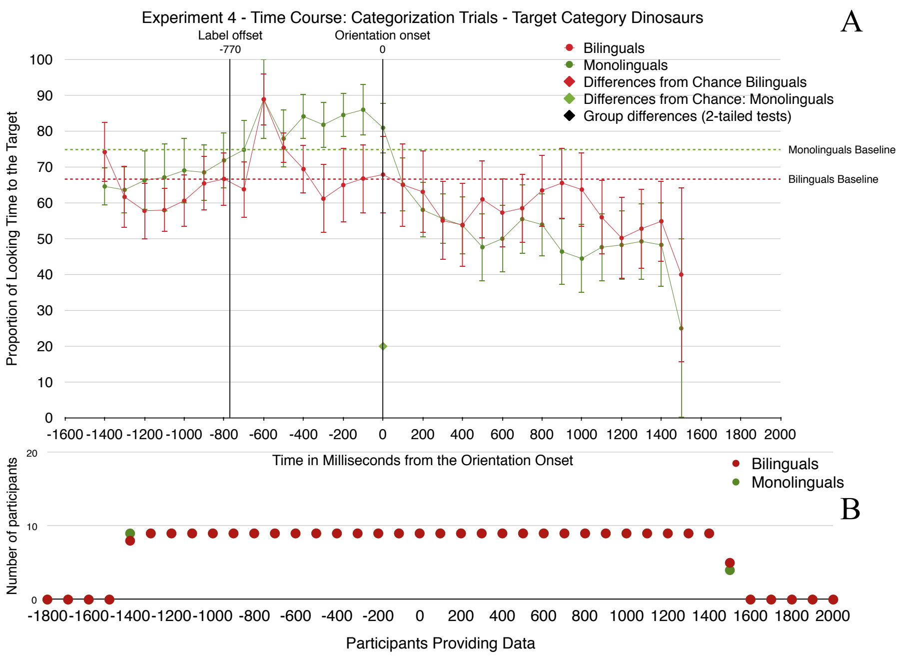


Figure 41. **A.** Symbols representing the PLTT across time in Categorization Trials of Experiment 4 for infants familiarized with the Target Category Dinosaur. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Infants who saw the dinosaur category reinforced during the familiarization trials started the trial orienting towards the target category. The mean PLTT during the Baseline period was above chance in the monolingual (Mean PLTT = 74.93; t (8) = 4.64; P = 0.0017) and bilingual groups (Mean PLTT = 66.61; t (8) = 4.87; P = 0.001). Both groups maintained high scores of the PLTT until the orientation onset and afterwards we observe a progressive drop of the PLTT. The time-course analysis did not reveal any time window in which bilinguals or monolinguals increased their PLTT relative to the corresponding baseline levels. No group differences where found (see Figure 41).

The results split by target category showed that both bilingual and monolingual infants that had seen the category fish reinforced in familiarization trials were not oriented towards the target category at the beginning of the trial. During the Post-Label period both showed an increase of the PLTT with respect to their baseline level. That increase was larger for the bilingual group. Bilingual infants familiarized with the dinosaur category started the categorization trials at higher PLTT scores and maintained high PLTT during the post-label period. Monolingual infants familiarized with the dinosaur category also started the trials at higher PLTT scores but decreased after label offset. Perhaps their high baseline starting point did not give room for any increase to happen.

To asses the progression of the label-category learning across the experiment, we conducted separate analyses for the first half of Categorization trials (Blocks 1-2) and the second halves (Blocks 3-4) of categorization trials. Results of the first half did not identify any time window in which bilinguals or monolinguals increased the PLTT after label. No group differences were found. In the second half of Categorization trials bilinguals did increase PLTT. The time window of the effects started 500 ms after the orientation onset and extended in time for 600 ms (that is, from 500 ms to 1100 ms after the orientation onset. For monolinguals, there were not signs of learning according to our criteria. We also found a short time window (300 ms) in which the dynamics of the two groups diverged (from 800 to 1000 ms after the orientation onset), but that window was too short to be reported as an effect (see Figure 42).

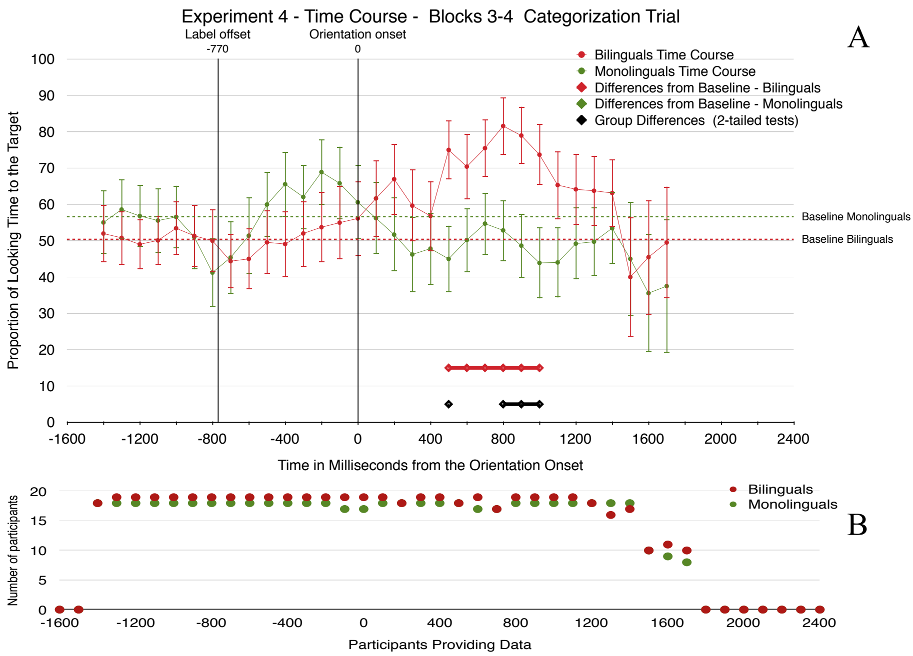


Figure 42. A. Symbols representing the PLTT across time in the second half of the Categorization Trials of Experiment 4. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. B. Plot representing the number of participants providing data in each time-bin.

These results suggest that the task remains demanding at 19 months, and effects only appear after several trial presentations.

Finally, a linear regression was calculated with Percentage of Exposure to a Second Language as a predictor and Change of PLTT between the Baseline and the Post-Label period in Categorization trials as a dependent variable. At 19 months, percentage of exposure to a second language did predict the Change in PLTT ($F(1,35) = 8.57$, $P = 0.006$, $R^2 = 19.7\%$). The higher the exposure to a second language, the stronger the PLTT increase in the Post-Label period (see Figure 43).

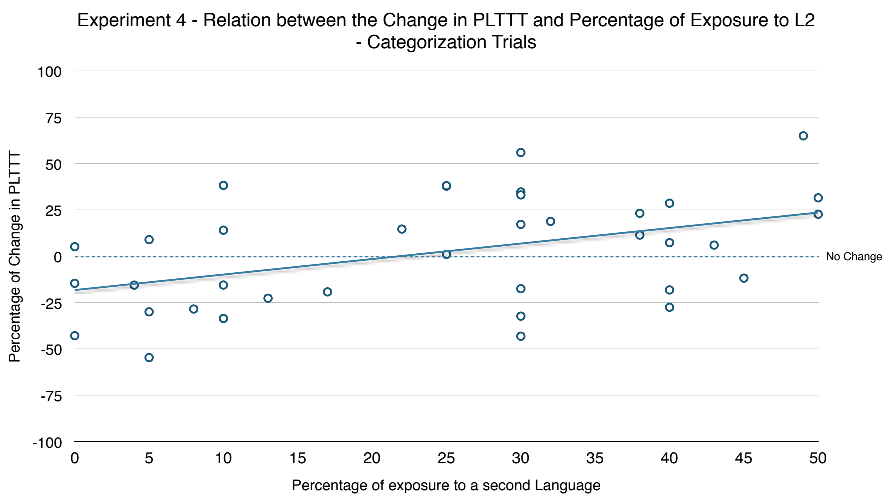


Figure 43. Relation between the Percentage of Exposure to a Second Language and the Change in PLTT in Categorization Trials of Experiment 4. The dashed line represents a null relation. The continuous line represents the regression line.

2.5.2.2.3. Incidental Mutual Exclusivity Trials

First, we conducted an overall analysis in which we extracted the average of the responses to the 4 trials of the IME condition.

A 2-way mixed ANOVA was run on the PLTT in the IME Trials with Time Window (Baseline/Post-Label) as a within – participants factor, Linguistic background (Bilinguals/Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background. Results did not detect any effect of Time Window (P = 0.71) or Linguistic Background (P = 0.70) on the PLTT. However, there was a marginal interaction of Time Window with Linguistic Background (P = 0.07) (see Figure 44).

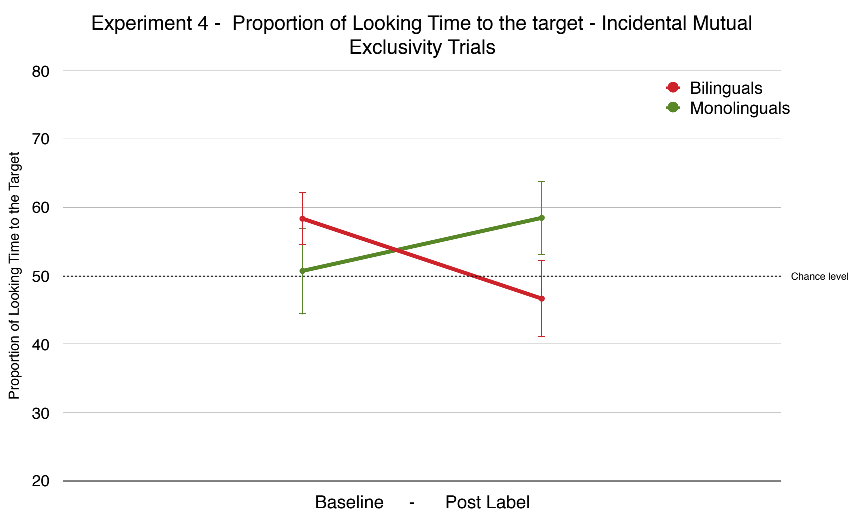


Figure 44. Mean PLTT during the Baseline and Post-Label period of the IME trials of Experiment 4. Bars represent SEs. * indicate effects at $p< 0.05$

Infants did not significantly modified their responses after listening to the label, but bilinguals, showed a trend towards reducing their

PLTT during the post label period -- a measure of success in our task.

We then conducted the same analysis including the independent variable Target Category. A three-way mixed-effect ANOVA was run on the PLTT in the IME Trials with Time Window (Baseline/ Post-Label) as a within – participants factor, Linguistic Background (Bilinguals/Monolinguals) and Category (Dinosaur/Fish) as a between – participants factor and Participants as a random factor nested in Linguistic background * Category. Results identified a main effect of Target Category as expected from to previous analyses. Also, a marginal interaction between Time window and Linguistic Background emerged ($P = 0.056$). Scheffé post hoc tests showed that bilinguals and monolinguals modified the PLTT scores during the Post-Label when they had seen the target category fish in familiarization trials. Interestingly, the change in performance was in the opposite directions for each group. Monolinguals clearly increased PLTT after the label ($P = 0.008$) while bilinguals reduced it ($P = 0.02$). There was also a triple interaction between Time Window, Target Category and Linguistic Background ($P = 0.004$). Scheffé post hoc tests showed that the effect was mainly driven by differences in the baseline scores of monolinguals between the two target categories, no other relevant effect were found (see Figure 45).

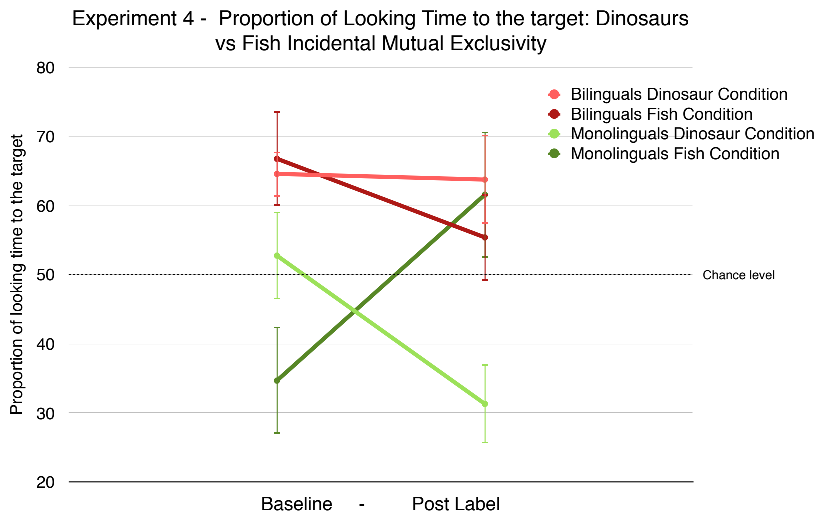


Figure 45. Mean PLTT during the Baseline and Post-Label period of the IME Trials of Experiment 4 split by Target Category. Bars represent SEs. * indicate effects at $p < 0.05$

Restricted analyses on blocks 1-2 and 3-4 of IME trials separately, showed that bilinguals presented a trend towards reducing the PLTT in the Post-Label period already during the first half of the experiment. A two-way mixed-effect ANOVA on the PLTT in Blocks 1-2 of IME trials with Time Window (Baseline/Post-Label) as a within – participants factor, Linguistic background (Bilinguals/ Monolinguals) as a between – participants factor and Participants as a random factor nested in Linguistic background. Results did not reveal any effects of Time Window (P = 0.91) or Linguistic Background (P = 0.71) on the PLTT. However, there was a marginal interaction between Time Window and Linguistic Background (P = 0.07). The direction of the trend suggested a reduction of the PLTT in the Post-Label period for bilinguals, indicating that the referential expectation may have appeared quite quickly. Monolinguals showed an opposite trend and tended to increase their PLTT after the label. No main effects or interactions were found when the same anova was conducted on Blocks 3-4 (see Figure 46).

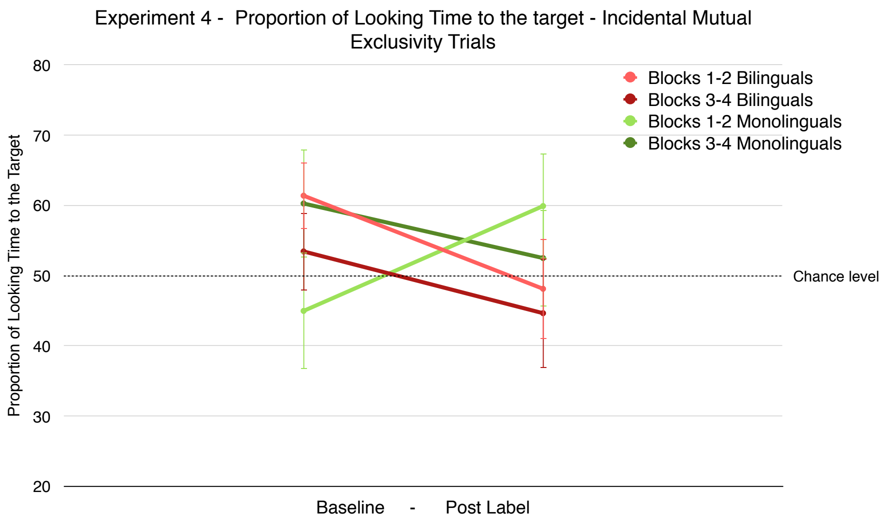


Figure 46. Mean PLTT during the Baseline and Post-Label period for the first and second half of the IME Trials of Experiment 4. Bars represent SEs. * indicate effects at $p < 0.05$

To better describe the effects revealed by the ANOVA analysis, we ran a time-course analysis of IME trials. Results revealed that bilinguals clearly reduced their PLTT after the label. The sequence went from 600 ms to 1700 ms after the orientation onset (1200 ms). No significant changes in monolinguals, nor group differences, were observed. These results show that bilinguals notably reduced their PLTT during the Post-Label period thus looking toward the unlabeled category (see Figure 47).

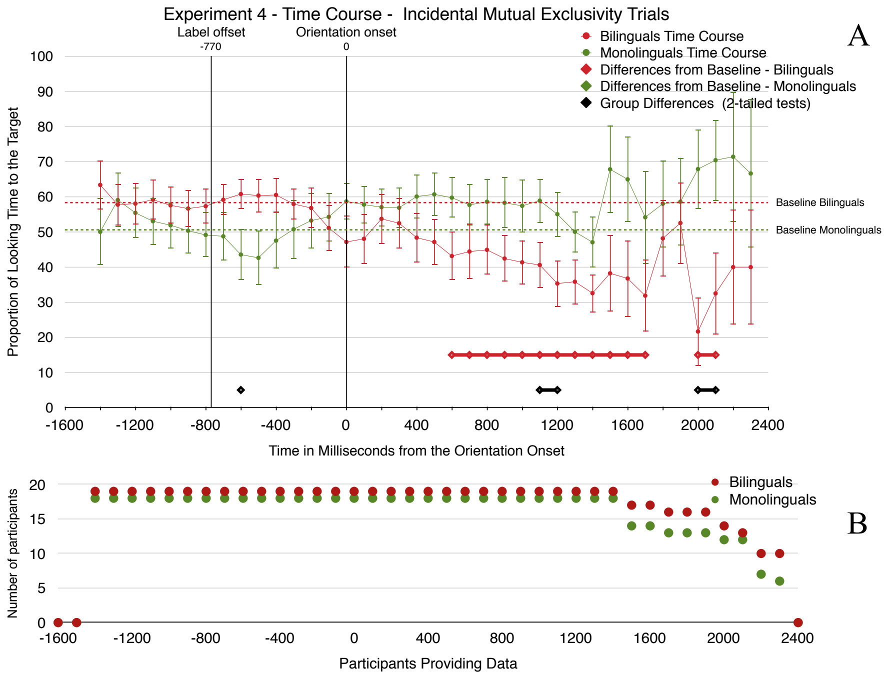
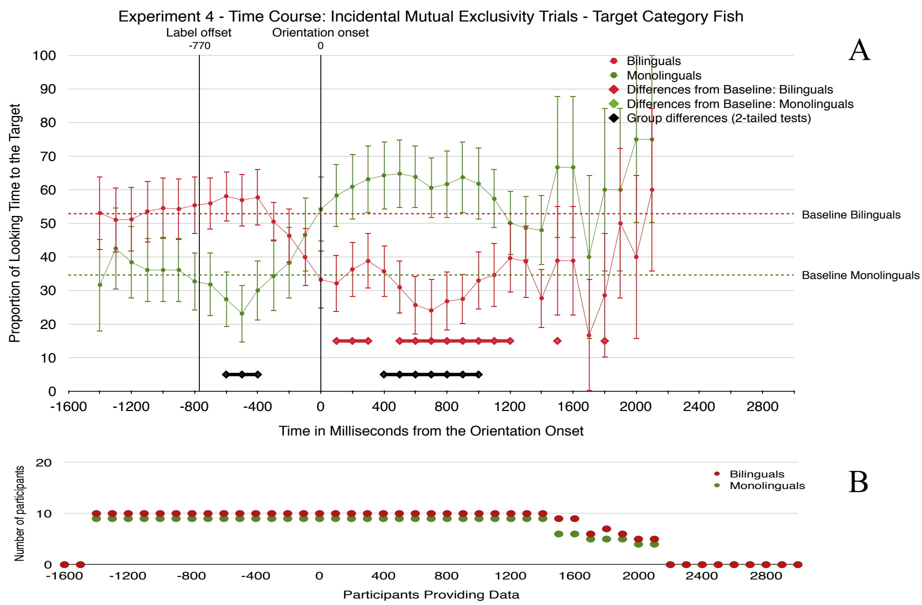


Figure 47. **A.** Symbols representing the PLTT across time in IME Trials of Experiment 4. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Then, separate time-course analyses were performed for each Target category type. The analysis of the infants that were familiarized with the target category fish revealed a long period of reduction in PLTT in bilinguals. It was an early effect, emerging around 500 after the orientation onset and extending for 800 ms (that is, from 500 ms to 1200 ms after the orientation onset). Also, group differences appeared starting 400 ms after the orientation onset (from 500 ms to 1000 ms after the orientation onset). That result seems to suggest that bilingual infants familiarized with the target category Fish clearly preferred to map the novel label with the category that was not labelled in previous trials. In monolinguals, the visual inspection of the temporal course seems to indicate that they presented the opposite pattern than bilinguals and they increased the PLTT after the label. The temporal course in the Post-Label period for monolinguals was reanalyzed using consecutive two-tail t-tests at $p < 0.05$ in order to confirm that possibility. The time-course analysis confirmed that effectively monolinguals increased the PLTT (the incorrect response in IME trials). A long sequence (1100 ms) was found starting 100 ms after the label offset (see Figure 48)

The time-course analysis of the infants that were familiarized with the target category dinosaur revealed a short time period (300 ms) in which bilingual participants decreased their PLTT, starting 1200 ms after the orientation onset. This period was too brief to reach significance according to our criteria. It was a small and late effect compared to the sequence found in bilinguals when the target category was Fish, but considering the large preference for Dinosaurs, no strong effect was expected. No effects appeared for monolinguals and there were no group differences. (see Figure 49)



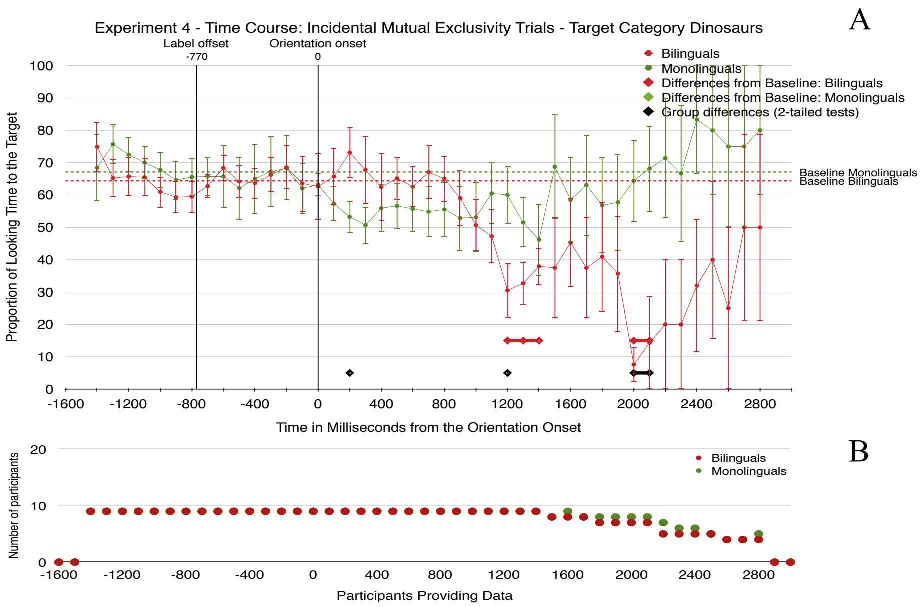


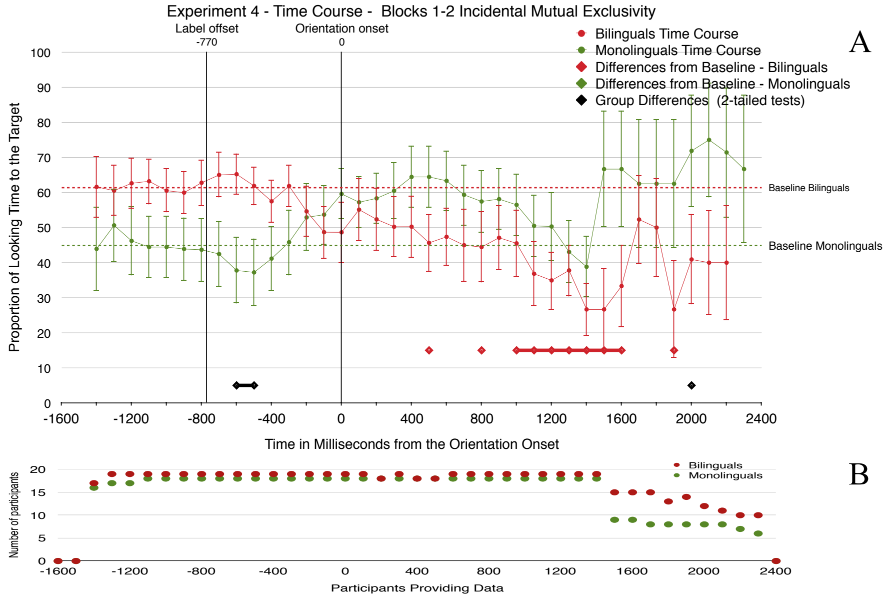
Figure 48-49. **A.** Symbols representing the PLTT across time in IME Trials of Experiment 4 for infants familiarized with theTarget Category Fish and Dinosaur. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Thus, the time-course analysis split by target category type showed that bilinguals tended to reduce their PLTT during the Post-Label period, although that tendency was stronger for infants familiarized to the target category fish. For monolinguals no clear pattern of responses emerged.

To explore the effects of presentation repetition on the learning of IME trials, a time-course analysis for the first (Blocks1-2) and second halves (Blocks 3-4) of IME trials were obtained. Essentially, the results were similar. In the first half, a clear trend towards the target category emerged for bilinguals, starting 1000 ms after the orientation onset. It was a late effect that started at the very end of the Post-Label window of interest, which is more evident when extending the window up to 1.5 s. Monolinguals, as seen in previous analyses, showed the opposite trend (see Figure 50).

The same profile, although much weaker, occurred in the second half of the experiment, with a short time interval of decrease of target looking (300 ms) at about the same time. No effects appeared in monolinguals (see Figure 51)

Altogether, the temporal analyses seem to indicate that bilinguals could successfully use the information extracted in Categorization trials to learn a second label-category pairing. Bilinguals switched the locus of attention from the category for which they had a name to the category that was not labeled in previous trials.



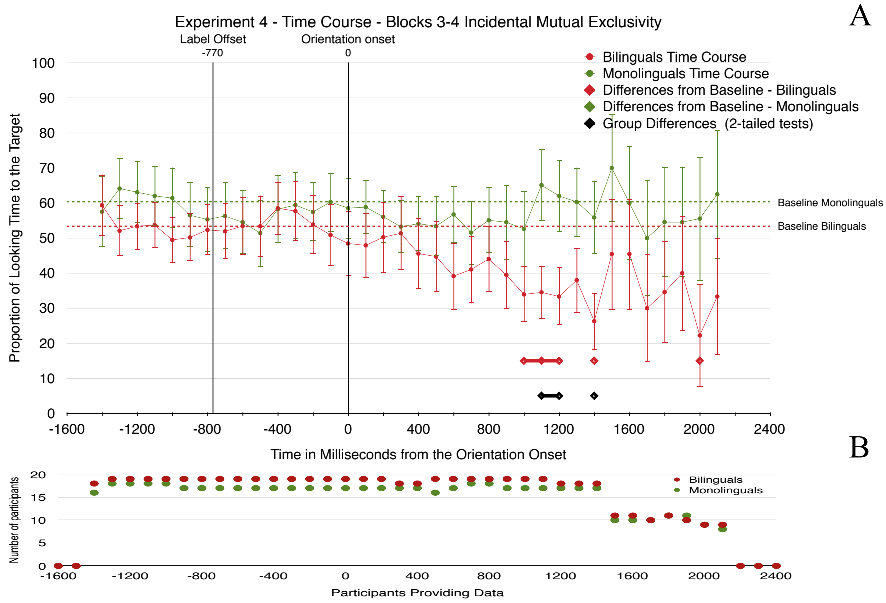


Figure 50-51. **A.** Symbols representing the PLTT across time in the first and second half of IME Trials of Experiment 4. Vertical lines represent the offset of the label and the onset of the orientation window. Horizontal lines indicate the baseline score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from baseline in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Interestingly, this response was observed in bilinguals already in the first half of IME trials, suggesting that deciding about the relation between the novel label and the novel category did not require much experience or repetitions. However, the dynamics of the response within trial showed a tendency to respond later than in Categorization trials. The late response during the trial, which nevertheless appeared right in the first IME trials of the experiment, suggests that IME trials engaged infants in a sort of local reasoning, requiring immediate, but more elaborate information processing during the trial -- a process that may delay the behavioral response.

Finally, a simple linear regression was calculated with Percentage of Exposure to a Second Language as a predictor and Change of PLTT between the baseline and the Post-Label period in IME trials as a dependent variable. Results showed that at 19 months of age, the Percentage of exposure to a second language is not a good predictor of performance in IME trials ($R^2 = 1.5$; $P = 0.47$).

2.5.3. Discussion

Nineteen-month-olds' response to our tasks gave us a picture which is difficult to interpret. In familiarization, infants behaved as expected if they had learned some relationship between the label and the category. In Categorization Trials, no main effects emerged if the nature of the Target Category was not taken into account, although a marginal interaction between Time Window and Linguistic background suggests language-specific differences. When Target category was considered, it appeared that infants who were familiarized with Dinosaurs maintained their preference after the orientation onset, whereas infants familiarized with the Fish switched to the target category after they listening to the familiar label.

The temporal course analysis was more informative about the dynamics of infants' responses, revealing effects that were difficult to spot from the overall analysis. Bilinguals presented a robust increase of the PLTT in the Post-Label window of the Categorization Trials, whereas no clear signs of learning appeared in monolinguals.

The block analysis showed that bilinguals tended to learn during the second half of the experiment. This result offers different interpretations. One possibility is that the task is very demanding at 19 months. However, in Experiment 2 we found that 8-month-old bilinguals could extract the category already during the first half of the experiment.

The missing piece to understand why 19-month-olds seem to learn a label-category pairing more slowly than 8-month-olds comes from the relationship between degree of bilingualism and Category learning: the percentage of exposure to a second language was a predictor of performance in categorization trials. This result might indicate that for 19-month-olds, learning moves from plain associative learning to more linguistic learning. Perhaps, at younger ages infants considered the relationship between the label and the target category as a simple association between a sound and a set of pictures. At 19 months, the label may acquire the role of a real potential word, thus leading infants to treat the associations as an occasion to acquire the relationship between a word and its abstract meaning. This additional linguistic component may add to the processing load of the task. The explanation is consistent with the fact that 19-month-olds reacted to the label-category association late during trials, and late across the experiment.

Also in the IME trials, no clear pattern emerged overall. We found a marginal interaction between Time Window and Linguistic Background that may indicate differences in the pattern of responses of bilinguals and monolinguals. Nevertheless, the time-course analysis clarified that language did have a role in our task at that age.

We saw that bilinguals significantly decreased the PLTT after the label onset in IME trials, that is, they oriented towards the novel target category for those trials. This result is very interesting for several reasons. First, it may suggest that 19-month-old infants can use recently acquired category-label pairings to boost new learning. Then, it may suggest that bilinguals rejected a second name for the category for which they just learned a label and preferred to map the novel label with the nameless category. This result runs counter the hypothesis that previous experience with two names for the same object would make bilinguals less adhered to use a ME strategy (Byers−Heinlein & Werker, 2009). This was not the case in our study. Because our task do not explicitly signal the language in which the labels were presented (labels were not embedded in any carrier sentence and were compatible to both languages) infants may have interpreted the two labels as belonging to the same language. Thus, the result might indicate that, although 19-month-old bilinguals can accept that two languages can give different names to the same object, within each language the 'one object, one name' principle still applies.

Furthermore, bilinguals needed few trial repetitions before orienting in IME trials. Bilinguals already oriented towards the novel target category already in the first half of the experiment. This suggests that coupling the novel word with the novel target category was a quite an immediate process, requiring little experience but engaging infants in an online reasoning *during the trial,* leading them to conclude that the novel label could not refer to the old category.

If we consider existing ME data and the current literature, the bilingual advantage on IME trials was unexpected. Previous studies showed that bilinguals are less inclined to use ME to disambiguate the meaning of new words for objects (e.g. Davidson & Tell, 2005; Byers−Heinlein & Werker, 2009; Houston−Price et al., 2010). (Although see Frank and Poulin-Dubois (2002) for results with older bilinguals). At the same time, our monolingual participants did not show clear signs of learning the label-category pairing or of solving the IME task. Even this result was unexpected on the basis of the current literature. Indeed, studies on monolingual infants at a similar age show that they use ME to learn novel words (e.g.Halberda, 2003; Byers−Heinlein & Werker, 2009; Markman et al., 2003). Thus, why did monolinguals not show signs of learning in IME trials?

We believe that the discrepant result depends on the differences between IME and the standard ME. One cannot succeed in IME without having previously extracted the category-label pairing in the familiarization trials. Nineteen-month-old monolinguals did not show clear signs of learning during the Categorization trials. If they did not learn to identify the target category members, they could not discard the target category member as a candidate. Then, we should ask, why did monolinguals not learn the word for the target category in familiarization? Possibly, more stability in the training sessions than a few reinforced repetitions interrupted by non-reinforced trials is needed. In short, our IME and fast Categorization trials require a certain dose of flexibility. Potentially, bilinguals are still more able to adapt to online changes of the learning problems that help them give appropriate responses to our tasks, especially to our IME task. Further experiments are needed to understand the exact behavior of monolinguals. A possible way to make progress on this issue is to compare bilinguals and monolinguals on a more 'classic' label-learning task for categories. This is the aim of our next experimental series.

3. EXPERIMENTS 5 TO 6: Slow Categorization and referent identification at 18 and 19 months

3.1. Introduction

The results of Experiment 4 were somewhat surprising. Apparently, 19-month-old bilinguals did what the literature suggest they would not do. Monolinguals did not do what the literature suggests they should do. We suggested that a possible explanation could be looked into the nature of the tasks we implemented, which require a strong component of flexibility and a particular attention to local relations between label and categories. While revealing, these tasks may not be word learning tasks after all. Word learning requires, yes, flexibility, but also stability: in the ever changing relations between sounds and objects, finding steady fixed points is as important as flexibility for learning words. We wanted to understand whether, in a paradigm that is not marked by the characteristics of fast fixation and fast reorientation that are essential parts of the previous experiments. In the following experiments, we adopted a classic approach to explore categorization in 19-month old infants. We adapted to our purposes the novelty-preference task used in Fulkerson and Waxman (2007) that has been widely used to explore the links between language and conceptual organization. Typically, in this procedure, infants are first either habituated or strongly familiarized to a visual category that can or cannot be paired with an auditory stimuli. In the test phase, participants are presented with the equivalent of forced choice trials, in which they can look at a new member of the familiar category or a member from a new category. A preference for the novel element of the test pair is understood as an indicator that infants extracted the category (Although other patterns of looking are possible depending on factors such as age the of participants or task complexityHunter & Ames, 1988). Previous studies using similar procedures showed that labels can guide referential expectations. By the end of the first year of life, consistent labelling induces categorization, while inconsistent labelling induces object individuation (Plunkett et al., 2008; Waxman & Braun, 2005). Similarly, when two labels correlated with visual features of the category to be learned, infants tend to form two categories, whereas when one label is assigned to all objects, infants tend to learn one single category (Plunkett et al., 2008).

In the following experiments, we explore the role of linguistic experience on the ability to map new labels to categories at 19 months. We will also explore how 18 and 19 months old bilingual and monolingual infants use this new knowledge to disambiguate the referent of a second label online. We focus on this age because in our previous experiments participants seemed to be sensitive to the role of labels, with the unusual results we reported. The much slower task that we implement offers more opportunity to realize a proper mapping between labels and categories, thus asking infants to solve a learning problem much more akin to word learning and much less demanding in terms of executive functions.

With that purpose, participants were tested with a procedure that consists of two phases. In a familiarization phase, infants are presented with different tokens of the same visual categories paired with one single constant label (Experiment 5), or with two visual categories consistently paired with two labels (Experiments 6). A test phase immediately follows the familiarization trials. It consists of three within-participants conditions, aimed at testing distinct aspects of the category representation potentially acquired during familiarization. In each test trial, infants see novel pairs of pictures, presented side by side, and listen to a label referring to one of the categories to which the images belong.

The first test type (*Broad Category Trials*), assesses if infants can recognize new members of each of the familiar category/ies when they are counterposed to pictures from novel categories. In the second type (*Narrow Categories Trials*), we evaluated the boundaries of the acquired category/ies. Finally, in the third type *(Incidental Mutual Exclusivity Trials),* infants are shown the same sort of visual contrast than in the Broad Category Trials (a new exemplar of the familiar category and a member of a new category) but they were presented with a new label. In this test trials the target category switches, and we measure infants' flexibility to use a recently acquired label-category pairing to reason about the possible referent of a new label in a paradigm which requires a more solidified acquisition of the previous target category.

Notice that in "standard" Novelty-preference paradigms (e.g. Fulkerson & Haaf, 2003), test trial images tend to be presented in isolation, not accompanied by any label. Nevertheless, we decided to present labels in order to induce different referential expectations. By adding a label we turned the task from a Novelty-preference procedure to a guided looking task and therefore the predictions about infants' behavior are different. In a Novelty-preference task, it is assumed that if linguistic stimuli facilitate category acquisition, infants will develop a preference for novel objects across the experiment. The underlying logic is that the category members will become familiar across presentations. Therefore, they will prefer to look at novel objects. Instead, we are interested in obtaining a dynamic measure of attention in response to a linguistic stimulus and record moment-to-moment information. For this reason, we are not particularly interested in overall preferences, but we seek for specific looking patterns that may reflect the underlying mental process that infants unfold during the task.

To sum up, the aim of Experiments 5 and 6 is to explore the role of linguistic background on 19-month-old infants' acquisition of labels for categories and of their representations, with a task that is closer to a word acquisition procedure as studied in the literature. We are particularly interested in comparing infants' behavior in this task with the unusual results that our previous task unveiled at the same age.

At the moment of submitting this dissertation, this work is in progress. Preliminary results are discussed, and should be considered in the light of the fact that it has been impossible to complete all experimental groups.

3.2. Experiment 5: A slow categorization task at 19 months of age

The aim of Experiment 5 is primarily to validate our adaptation of a familiarization/preferential-looking categorization task and try to replicate previous results showing that labelling facilitates categorization (Fulkerson & Waxman, 2007). Second, we are interested in exploring the influence of linguistic experience in a slow categorization task. Specifically, we ask whether our previous findings with a 'fast labeling' procedure—a bilingual advantage in categorization and reorientation to novel labels—still hold with the current task.

We tested 19-month-old bilingual and monolingual on their ability to acquire new names for categories (Broad Categorization Tests) and to discriminate the target category from close competitors (Narrow Categories Test). Finally, we asked whether participants could infer the referent of a new label online in an Incidental Mutual Exclusivity task (IME trials).

Because before 19 months of age categorization is well documented, we expect that both bilinguals and monolinguals should succeed in learning a novel word for a category under the conditions of presentation of the current task. However, a more articulated pattern of responses may emerge in the IME trials. Bilinguals, accustomed to the presence of translation equivalents, my accept that categories can be associated with different labels. Monolinguals, on the other hand may maintain the constraint that one entity has only one word. If this is the case, monolinguals may outperform bilinguals in the IME task.

3.2.1. Materials and methods

3.2.1.1 Participants

Thirty-four full term 19-month-old infants were retained for analysis: 21 Monolinguals (14 girls. Mean age: 18;23, Range: 17;26 – 19;29) and 13 Bilinguals (6 girls. Mean age: 19;23, Range 18;05 – 20;09). All participants were healthy and free from birth complications according to the parent's report.

Eighteen additional infants (11 monolinguals and 7 bilinguals) were tested but not included in the analysis because they cried or refused to be seated (8), they were inattentive (2), there was parental interference (2), or the equipment failed (6).

The total rejection rate was 35%. We want to highlight that in this experiment some participants were tested after running another task previously. The previous task had no relation with the current experiment, but it may have depleted some of the attentional resources needed to complete a long procedure such as the one used here.

The bilingualism definition and the recruitment procedure were identical to that of Experiment 1.

3.2.1.2. Materials

3.2.1.2.1. Stimuli

We constructed 28 monochromatic drawings on a grey background, representing the category Dinosaur, split into two subcategories: 14 exemplars of 4-legged dinosaurs and 14 exemplars of 2-legged dinosaurs. The pictures of 2-legged dinosaur exemplars were digitally modified so as to increase the contrast between sub-categories and to avoid confounding with number of legs. Another additional 4 images were created as novel contrasting categories for the test phase (fish, bird, insect and turtle). Some of the images had been used in the previous experimental series.

The images used during the familiarization phase varied in color to make the exemplars more distinguishable. The images used during the test phase were all matched in color, in order to minimize preferences for one picture over the other not induced by the experimental task.

The images were edited with the Graphic Converter 7.6.2 software and later prepared as animated slides with Keynote 5.0.5 (Apple iWork '09) software. Finally, the slides were exported as QuickTime videos of 400 x 400 pixels, at 24fps, and MPEG-4 video format. As in the previous experimental series, in the familiarization phase the images could expand and contract 20% of their size when played.

The auditory stimuli used were the same as in Experiment 1. In this procedure the labels (*Mapu* or *Doti*) could be presented in isolation, as in Experiment 1, or could be embedded in a carrier sentence. We recorded 4 carrier sentences of the type “Oh! Look at the *x*” , where *x* stands for the label. These sentences were used in the familiarization phase to reinforce the language context in which the task was given. Each token was crossed with label (2) and language (2) yielding 16 recordings. Files were matched in duration across languages and labels.

Finally we recorded 17 short clips (8 videos in Catalan, 8 videos in Spanish, and a silent video of a smiling woman). The actress spoke in a motherese style and uttered child-directed sentences (“You are doing great” or “Hey baby! Do you want to see another movie?”). Videos were edited and matched for duration with Adobe Premiere CS5 software to create almost identical versions in each language.

3.2.1.2.2. Apparatus

The experimental setup was identical to the one described in Experiment 1.

3.2.1.3. Procedure

We followed the same calibration procedure used in previous experiments (see Experiment 1). Infants were randomly assigned to one out of eight conditions: Visual category (2 or 4 legged Dinosaurs) x Label (Mapu or Doti) x Randomization (2 Test Trial Order). Importantly, once established the label-category pairing was kept constant during the Familiarization Phase. Features other than category membership or label that could induce categorization were controlled across lists: Side of Appearance of the target picture (Left or Right) and Head Orientation of the drawings (Inside or Outside). We also controlled that the language in which the stimuli was presented corresponded to the dominant language in the infant's environment.

The task started immediately after the calibration. All trials began with a central attractor (400 x 400 pixels), a colorful rotating cross, to prevent orientation towards either any of the target ports before the appearance of the images. When the infant fixated to the attractor, the experimenter triggered the trial.

The procedure included a Familiarization phase (8 trials) followed by a Test phase (6 trials). The total duration of the task was around 5 minutes (see Figure 52).

Familiarization Phase

All infants were shown 8 different tokens of the target subcategory (either a two-legged or a four-legged dinosaur) presented sequentially in a 400 x 400 px port for 16.000 ms each. The presentation side (Left or Right) was randomly assigned with the constraint that no more than three pictures could be presented in the same location, so as to prevent infants from associating the appearance side with the visual category. The presentation order of the images in familiarization trials was random.

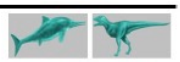
Condition	Auditory Stimuli	Label	Referent	Visual Stimuli
Familiarization	"Ostres! Això és un MAPU" x 3	Label 1	2 legs Dinosaur	
Familiarization	"Mira! un MAPU!" x3	Label 1	2 legs Dinosaur	
Familiarization	"Oh! Veus el MAPU?" x 3	Label 1	2 legs Dinosaur	
Familiarization	"Apa! És molt bonic aquest MAPU!" x 3	Label 1	2 legs Dinosaur	
Familiarization	"Ostres! Això és un MAPU" x 3	Label 1	2 legs Dinosaur	
Familiarization	"Mira! Un MAPU!" x 3	Label 1	2 legs Dinosaur	
Familiarization	"Oh! Veus el MAPU?" x 3	Label 1	2 legs Dinosaur	
Familiarization	"Apa! És molt bonic aquest MAPU!" x 3	Label 1	2 legs Dinosaur	
Broad Category Condition	MAPU	Label 1	2 legs Dinosaur	
Narrow Categories Condition	MAPU	Label 1	2 legs Dinosaur	
Incidental Mutual Exclusivity Condition	DOTI	Label 2	Novel Category	

Figure 52. Experiment 5 design. Infants were presented with 14 trials organized in 8 Familiarization Trials and 6 Test Trials (Two test trials per condition).

Once an image was presented it started to loom automatically, expanding and contracting 20% of its size. The movement was aimed at maintaining infants' attention. Infants listened three tokens of the label on each trial. The first presentation was always a phrase that contained the label in the last position. In the second and third presentations, infants listened the label in isolation. The onset of the three label presentations was pseudo-random, varying its onset between one and two seconds. That way, infants could not anticipate the exact onset of a label. Labels could not last longer than the trial length (see Figure 53).

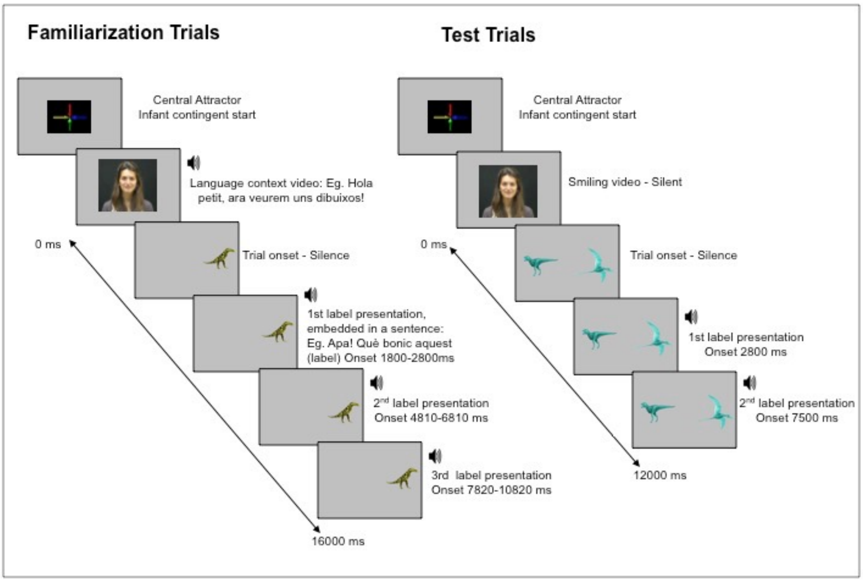


Figure 53. Trial structure of the familiarization and test trials in Experiment 5.

On trials 1 and 5 the attractor was followed by a short video of a young woman talking in infant-direct speech in the main language of the infant. The clip was randomly selected out of the 8 pre-recorded videos. Once a movie had been played it was not shown again. The aim of these videos was to reinforce the linguistic context of the task and to refresh infants' attention.

Test Phase

The test phase was comprised of 6 trials, two of each test type.

In the *Broad Category Trials,* participants were presented with a novel exemplar of the target (sub)category induced during familiarization, coupled with a member of a novel category (E.g. 2-leg dinosaur vs. a bird). Infants listened to the same label that had been paired with the familiar pictures during familiarization. The role of Broad Category Trials is to asses learning of the target category-label pairing induced during familiarization.

In the *Narrow Category Trials,* infants were presented with a new member of the familiar category (E.g. A 2-legged Dinosaur) and a member from a new subcategory of dinosaurs (E.g. a 4-legged Dinosaur). Again, infants listened to the label that had been assigned to the familiar category in the familiarization phase. These trials asses the specificity of the label-category pairing. If infants formed a narrow category (that is a subcategory of dinosaurs) based on the exemplars seen during the task, then these trials should elicit orientations towards the familiarized target category. Instead, if the nature of the label-category association is broad, infants may not have any preference among the two pictures (excluding novelty effects for the sake of the argument), because for them the label refers to *Dinosaurs* and not to *Two-legged dinosaurs* or *Four-legged dinosaurs*.

In the *IME Trials,* infants were presented with a visual contrast as in the Broad Categorization Trials (a new familiar member and a picture from a novel category). However, infants listened to a novel label that had not been paired with any category in previous presentations. The rationale for these trials is exactly as in Experiments 2-4.

In all test trials, two static pictures were presented side by side in ports of 400 x 400 px separated by 1120px for 12.000 ms each. Test images did not loom, even if an infant looked at the target image.

Therefore, infants did not receive any feedback. In every trial, the label was played twice in isolation, meaning not embedded in any carrier sentence. The onset of the first token was 2800 ms after the picture appeared onscreen. The onset of the second label was 7500 ms after the appearance of the pictures (see Figure 53).

Two trials of each condition type were presented across the test phase. Half of the participants saw two cycles of a Categorization trials followed by a Narrow Categorization trial and an IME Trial, whereas the other half saw the same sequence in reversed order. That is, they started with an IME trial.

3.2.2. Results

3.2.2.1. Scoring

To ensure that infants had the opportunity to encode the images, participants who looked at the images for less than the 50% of the total potential time during familiarization were excluded from the analysis. No infants had to be removed from Experiment 5 using that exclusion criterion.

The eye fixations were sampled at 60Hz with the Eye Tracker. In the Familiarization trials, each datapoint was classified as *Looking to the Target* or *Not-looking to the target*. In test trials, data points were classified as *Looking to the Target, Looking to the Non-Target* or *Looking to None of the Images*. The target object was defined as the object belonging to the category that had been associated with the label in familiarization trials.

As in the previous experiments, to calculate infants' responses, we only considered the gazes that fell into the pre-defined Regions of Interest (ROIs). ROIs coincided with the experimental ports where the stimuli were presented.

Proportion of looking time to the target (PLTT) was used as a measure of Categorization. In the familiarization trials it was calculated as follows:

$$\textbf{PLTT} = \frac{\text{Data Points to the target}}{\text{Potential datapoints}} * 100$$

In the test trials, it was calculated with the following formula:

$$\textbf{PLTT} = \frac{\text{Data Points to the target}}{\text{(Datapoints to the target + Datapoints to the non-target)}} * 100$$

The silent period of 2800 ms before to the onset of the first Label presentation (*Pre-Label Window*) was defined as our Baseline window. We analyzed the labeling effects by comparing participants' oculomotor behavior after the two label presentations *(Post-Label 1 and Post-Label 2 Windows*) relative to the Pre-Label Window. Each Post-Label window had a duration of 2500 ms.

For each participant, we calculated the PLTT during the Pre-Label period and each of the Post-Label Windows. Scores above 50% indicate a preference for the category that had been associated with the label in the familiarization trials (familiarity preference). Scores below 50% indicate a novelty preference in favor of the visual category not shown in the familiarization trials. We also computed the PLTT *across time*, dividing the time period of interest into time bins of approximately 100 ms each (6 data points). (see §2.2.2.1. for a detailed description of how to analyze the PLTT across time)

Following (Guthrie & Buchwald, 1991), for a sample of N= 15 participants, a window length of 25 time bins (or 2500 ms), assuming the highest degree of data autocorrelation, and a significance level of $P < 0.05$, 6 or more consecutive significant *t*-test comparisons indicate a reportable difference in the looking time pattern. Given that we expect a direction of effects, we used one-tail ed *t*-tests in order to compare the experimental conditions against their Pre-Label periods. However, because we have no firm prediction about the direction of the effect when comparing bilinguals and monolinguals, we used two-tailed *t*-tests for these comparisons.

The target category was the category member presented in the Familiarization Trials for all the trials, except for the IME trials, in which the target category was the alternative category. Thus, success in the task would be detected by an increase in PLTT in all the trial types except the IME trials, where it would amount to a reduction of PLTT.

3.2.2.2. Results

Preliminary results on the Familiarization Trials did not detected any effect of sex (Girls or Boys)(P = 0.30), Label (Doti or Mapu)(P

= 0.96) or Familiar Category (2-legged or 4-legged Dinosaurs)(P = 0.44) on infants' PLTT scores. Therefore Trials were collapsed across these conditions. To correct for baseline preferences for one category, in the test trials we compared infant's performance, not against chance, but against their Pre-Label preference scores.

3.2.2.2.1. Familiarization Trials

The results reported here include the averaged responses to all the Familiarization trials collapsed.

For each participant, an averaged score for each time window of interest was obtained. We first ran a two-way mixed ANOVA on the PLTT in Familiarization trials with Time Window (Pre-Label/Post-Label 1/Post-Label 2/Post-Label 3) as a within–participants factor, Linguistic background (Bilinguals/Monolinguals) as a between–participants factor and Participants as a random factor nested in Linguistic background. The results showed a main effect of Time Window ($M\ PLTT_{Pre\text{-}Label}$ = 66.17, $M\ PLTT_{Post\text{-}Label\ 1}$ = 79.46, $M\ PLTT_{Post\text{-}Label\ 2}$ = 71.68, $M\ PLTT_{Post\text{-}Label\ 3}$ = 62.61; $F(1,96) = 14.92$; $P < 0.0001$). Post hoc Scheffé tests revealed that infants increased the PLTT between the Pre-Label and the Post-Label Period 1($P < 0.0001$). However they reduced the PLTT between from Post-Label 1 and the Post-Label 2 Period ($P = 0.04$).

There was no effect of linguistic background ($P = 0.79$), nor an interaction between Time Window and Linguistic background ($P = 0.41$). Thus, results showed that both bilinguals and monolinguals increased their PLTT during the period following the label presentations. The increase occurred especially after the first presentation.

Then, a time-course analysis on the PLTT of the familiarization trials was conducted to better characterize the response dynamics. (see Figure 54).

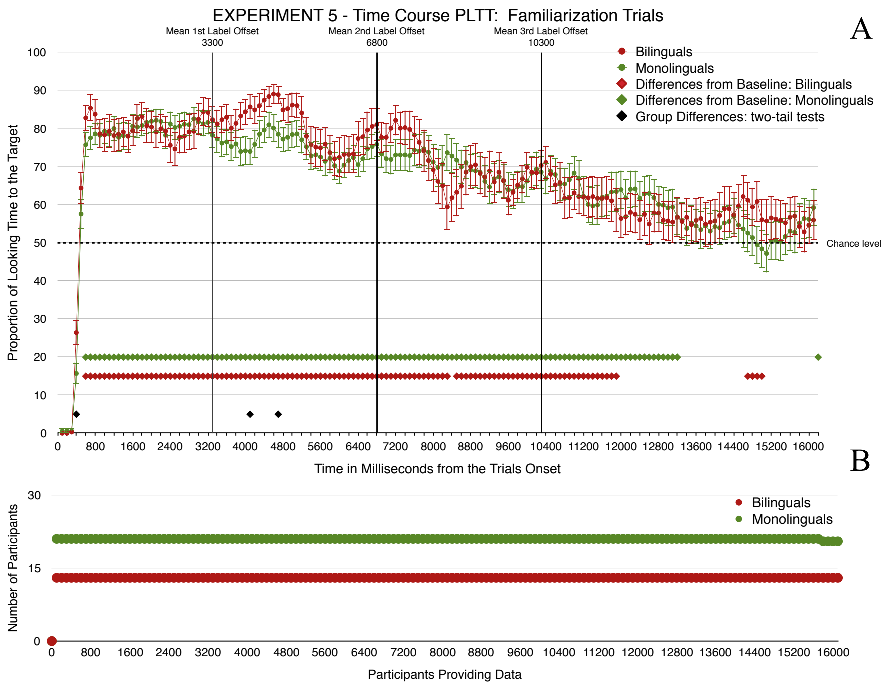


Figure 54. **A**. Symbols represent the PLTT across time in Familiarization Trials in Experiment 5. Vertical lines represent the Mean offsets of the three label presentations. The Horizontal line indicates the Chance level. Also are shown symbols representing the sequences of significant one-tail t-test at $p< 0.05$ different from Chance level for the bilingual and monolingual groups and the significant sequences of two-tail t-test at $p< 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

We found two long time windows in which participants performed above chance. Results revealed a significant sequence of one-tailed t-tests at $p < 0.05$ for monolinguals that started around 500 ms after the trial onset (Between 500 and 13100 ms after the trial onset). In bilinguals, the sequence also started at 500 ms but it was shorter (Between 500 ms and 11800 ms after the trial onset). We did not find significant differences between groups

The results of the two analyses showed that bilinguals and monolingual infants attended to the images similarly. They presented the highest PLTT scores during the Post-Label 1 period and their PLTT score decreased as the trial time passed.

3.2.2.2.2. Broad Category Trials

The results reported here include the averaged responses to the two Broad Category Trials collapsed.

Because in familiarization trials it was found that infants' attention tend to decrease across label repetitions we decided to restrict all our analyses to the first label presentation.

For each participant, an averaged score for the Pre-Label and the Post-Label 1 period was obtained. We ran a two-way mixed ANOVA on the PLTT in Broad Category Trials with Time Window (Pre-Label/Post-Label) as a within–participants factor, Linguistic background (Bilinguals/Monolinguals) as a between–participants factor and Participants as a random factor nested in Linguistic background. The results did not show any main effects of Time

Window (P = 0.13) or Linguistic Background (P = 0.47). There was no interaction between Time Window and Linguistic Background (P = 0.51) (see Figure 55).

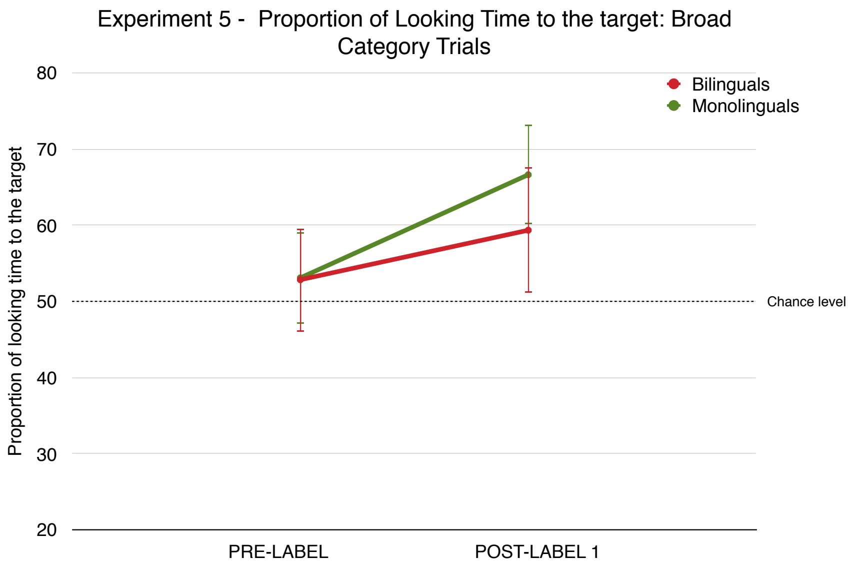


Figure 55. Mean PLTT during the Pre-Label and Post-Label period for the Broad Category Trials of Experiment 5. Bars represent SEs. * indicate effects at p <0.05 Bars represent SEs. * indicates effects at p< 0.05

No clear signs of learning the label-category pairing were found. The analysis did not reveal any increase of the PLTT following the first label presentation for any of the groups. Still, the visual inspection of the data suggests that there was a trend towards increasing the looking time to the familiar picture.

The time-course analysis of the PLTT could be more informative. Each time-bin in the 2.5s of the Post-Label 1 time window of interest was compared against the Mean PLTT score for the Pre-

Label period. An independent baseline score for the monolingual and the bilingual group was obtained.

In the Post-Label 1 period, a time window of 1000 ms emerged in which monolinguals clearly increased the PLTT above the baseline. The string started 1500 ms after the first label offset (that is, from 4300 to 5200 ms from the trial onset). No effects were found in bilinguals. At the same time, there were no group differences, suggesting that bilinguals were not very different from the monolingual infants. It should be noted, however, that the bilingual group we could test was quite small and amounts to only half of the sample needed to counterbalance the factors properly. Thus, any conclusion from the current data has to be taken with great care. (see Figure 56).

The results of the time-course analysis suggest that monolinguals successfully extracted the label-category pairing and were able to recognize new tokens of the familiar category. In bilinguals, we did not find clear signs of learning. although their scores did not differ from that of monolinguals in the region in which effects for that group were found.

The inspection of the temporal course also showed that both groups started with a strong familiarity preference during baseline. Consistent with Byers−Heinlein and Werker (2009), we averaged the period before the word onset to compute our baseline.

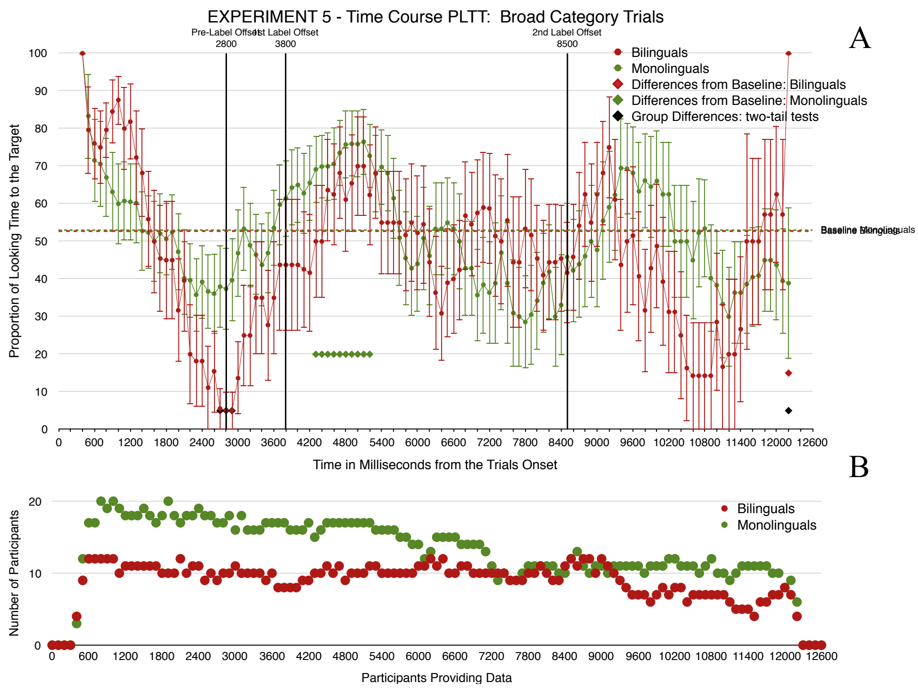


Figure 56. **A.** Symbols representing the PLTT across time in the Broad Category Trials of Experiment 5. Vertical lines represent the offsets of the two label presentations. Horizontal lines indicate the Pre-Label reference score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from the Pre-Label score in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

However, clearly this period is in itself dynamic, with the strong preference for the familiar category yielding to inspection of the competitor picture, before word offset. The selection of another baseline criterion, perhaps starting at the moment when the preference for the familiar picture withers down, would give a different picture of infants' behavior. We have no guidance as to how to proceed here, and thus we adopt the conservative stand of averaging the pre-label period over a very dynamic behavior. Further research is needed to settle on more objective criteria to take such decisions, which often make it difficult to understand the results reported in the literature.

3.2.2.2.3. Narrow Categories Trials

The results reported here include the averaged responses to the two Narrow Categories Trials.

We ran a two-way mixed ANOVA on the PLTT in Narrow Categories Trials with Time Window (Pre-Label/Post-Label 1) as a within–participants factor, Linguistic background (Bilinguals/ Monolinguals) as a between–participants factor and Participant as a random factor nested in Linguistic background. The results did not show any main effect of Time Window (P = 0.86) or Linguistic Background (P = 0.55). There was no interaction between Time Window and Linguistic Background (P = 0.30) (see Figure 57).

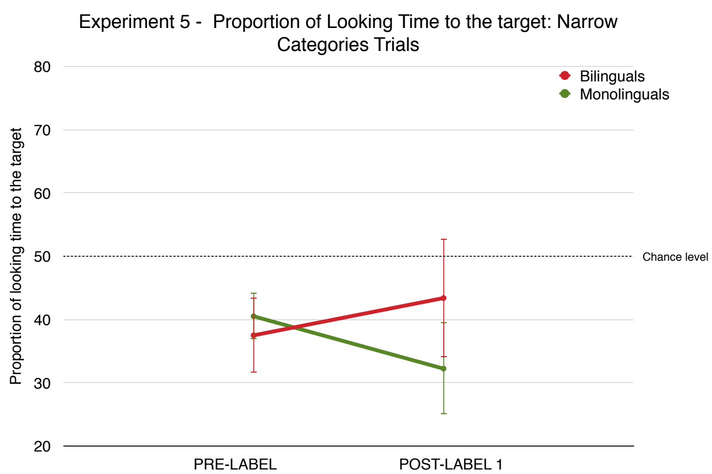


Figure 57. Mean PLTT during the Pre-Label and Post-Label period for the Narrow Categories Trials of Experiment 5. Bars represent SEs. * indicate effects at $p < 0.05$

Then we conducted a time-course analysis of the PLTT in the Narrow Categories Trials. We did not find any time window in which bilinguals or monolinguals increased the PLTT above the baseline reference during the Post-Label 1 period. No group differences were found (see Figure 58).

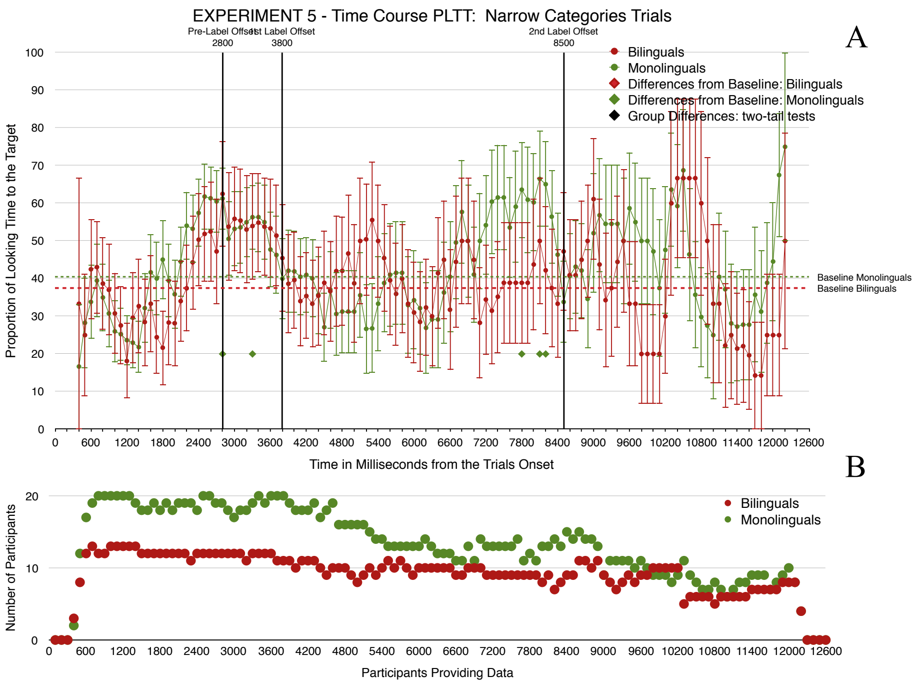


Figure 58. **A.** Symbols representing the PLTT across time in the Narrow Categories Trials Trials of Experiment 5. Vertical lines represent the offsets of the two label presentations. Horizontal lines indicate the Pre-Label reference score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from the Pre-Label score in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

These results suggest that infants did not differentially respond when two similar referential candidates were presented. It is possible that they did not encode the specific features that make the category members parts of different subcategories, or that they were simply confused by the test. In (Plunkett et al., 2008), 10-month-old infants formed two subcategories when they were presented with two labels, one for each subcategory that was consistently paired to all of its members. It is possible that explicit guidance by means of paring a label to each subcategory it is necessary in order for infants to attend to the relevant features that define that subcategory. In Experiment 6, we will try to address this question.

3.2.2.2.3. Incidental Mutual Exclusivity Trials

The results reported here include the averaged responses to the two IME Trials collapsed.

We ran a two-way mixed ANOVA on the PLTT in IME Trials with Time Window (Pre-Label/Post-Label 1) as a within–participants factor, Linguistic background (Bilinguals/Monolinguals) as a between–participants factor and Participants as a random factor nested in Linguistic background. There was no effect of Time Window (P = 0.23) or Linguistic Background (P = 0.31), nor interactions (P = 0.95) (see Figure 59).

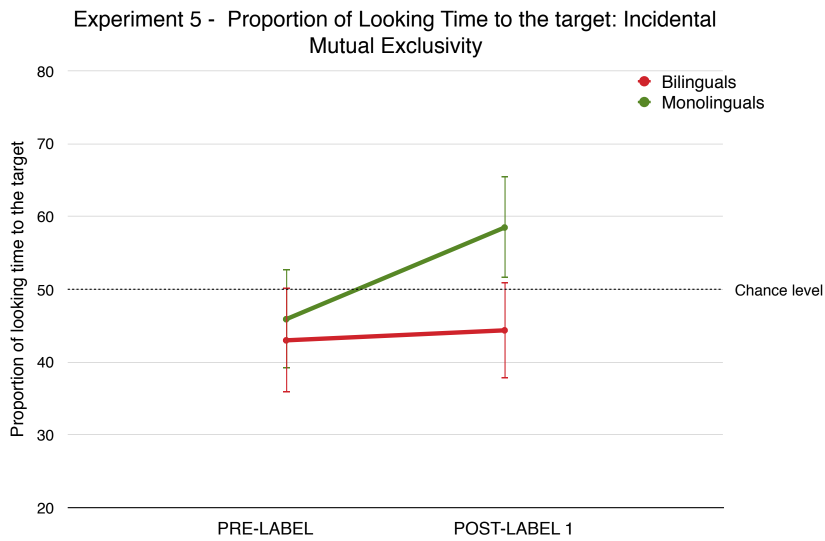


Figure 59. Mean PLTT during the Pre-Label and Post-Label period for the Incidental Mutual Exclusivity Trials of Experiment 5. Bars represent SEs. *

Finally, we performed a time-course analysis of the PLTT in IME trials. The analysis identified a stretch in which bilinguals reduced their PLTT compared to their baseline score (our measure of success in IME trials) after listening to the label presentation. The time window started 2400 ms after the label presentation and lasted for 600 ms. It was a late effect; the sequence stretched from 5200 to 5700 ms from the trial onset. No such effect was found in monolinguals. Group differences appeared among bilinguals and monolinguals in a time window of 700 ms that started 2300 ms after the first label presentation (That is, from 5100 ms to 5700 ms after the trials onset).

A close inspection of the temporal course suggests that monolinguals might be showing the opposite pattern than bilinguals. They direct towards the familiar category, just as in the Categorization trials. In order to confirm this trend, we notice that two-tailed t-tests in that period would statistically confirm that bilinguals directed their gazes towards the familiar category after the second label presentation. The sequence started 2400 ms after the second label presentation and lasted 900 ms (that is, from 10900 ms to 11700 ms after the trial onset) (see Figure 60).

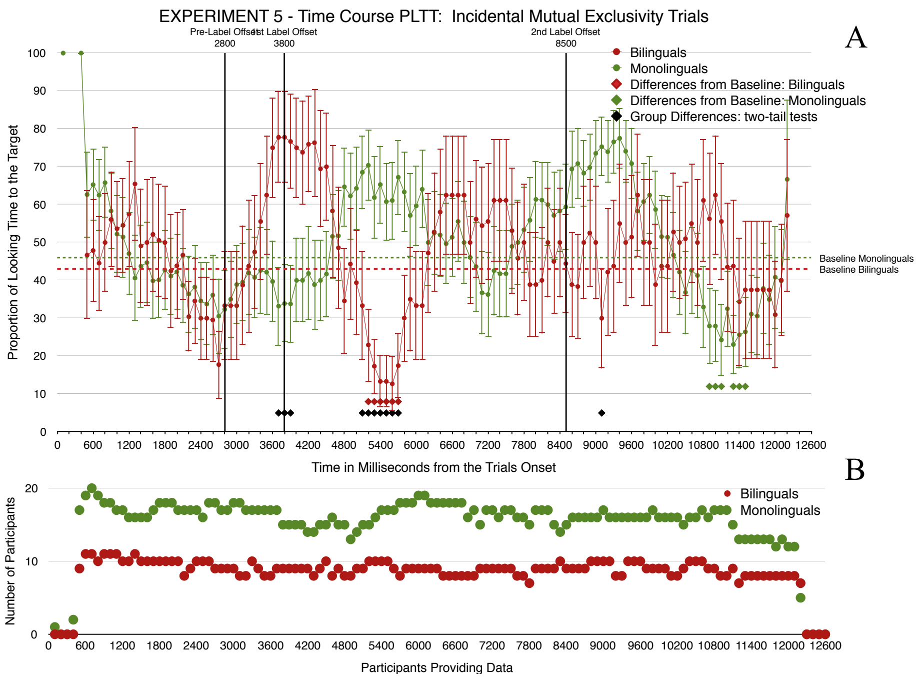


Figure 60. **A.** Symbols representing the PLTT across time in the IME Trials of Experiment 5. Vertical lines represent the offsets of the two label presentations. Horizontal lines indicate the Pre-Label reference score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from the Pre-Label score in monolinguals and bilinguals and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Therefore, although overall analyses revealed no effect, an inspection of the time-course of infants' gaze shows that bilinguals and monolinguals diverged after the first label presentation. Bilinguals showed a strong preference for the novel category after listening to a new name, suggesting that perhaps they associated the new label to the novel category. Even more interestingly, the dynamics of bilingual's gaze indicates some sort of doble-checking process: while listening to the label and afterwards, bilinguals first direct their gazes at the familiar category, and then clearly orient towards the novel category. This gaze pattern is consistent with the disjunctive syllogism strategy described in (Halberda, 2003). Monolinguals, instead, only directed their attention towards the familiar category, without any sign of double-checking or change in strategy across time. In both groups, the response came late in time, almost at the end of the predefined temporal regions of interest that we selected for analysis.

3.2.3. Discussion

We caution again that the data of Experiment 5 are preliminary results, and interpretations are only tentative.

Experiment 5 showed that the 19-months-old monolinguals tested with a “slow” categorization task successfully extracted the label-category pairing during the familiarization phase. When they were presented with two possible referents for the familiar label in the Broad Category Trials, monolinguals increased their attention to the image of the familiar category. This result replicates the previous findings in monolinguals supporting facilitatory effects of labelling in object categorization (e.g. Balaban & Waxman, 1997; Fulkerson & Waxman, 2007).

In contrast, in Experiment 4 we saw that monolingual infants of the same age did not show learning when they were tested on a fast categorization task that required rapid orientations and responses. This pattern may suggest that when monolinguals are tested with a less demanding procedure in terms of flexibility, they can learn a label-category pairing without any problem. It is also possible that a less flexible response might even be beneficial for some particular aspects of the slow categorization task such as the encoding of information relevant for word learning.

Bilinguals in Experiment 5 showed no clear signs of learning the label-category pairing. Interestingly, when bilinguals of the same age were tested on a fast categorization task in Experiment, they showed strong signs of learning the label-category contingency. Following the same line of thought, bilinguals in Experiment 5 would not get any particular advantage from being flexible when the task taps onto another kind of ability. Yet, it is also possible that the small sample size and a high initial familiarity preference may overshadow a potential result for the bilingual group. We notice that this conclusion is also dependent on the baseline used to perform the analysis.

To sum up, it is possible that our "fast" and "slow" categorization tasks are tapping into different abilities.

In the Narrow Categories Condition, none of the groups were found to be able to recognize the familiar referent when the contrasting image belonged to a close visual category. It might be that the category representation that infants formed through the task did not contain sufficient detail as to allow the discrimination. We do not argue that infants cannot perceive the difference between 2 vs 4 legs. Rather, if their attention is not explicitly directed towards those differential features of subcategories, then perhaps they do not encode them as defining traits to fix the label-reference relation. Perhaps, because the contrasting category belonged to the same basic category as the target and looked similar, infants generalized the familiar label to the full category of dinosaurs.

Finally, in the IME trials we found an interesting pattern. The analysis of the temporal course revealed that bilinguals tended to look to the novel category when a new label was presented. Monolinguals instead showed no such trend, and increased the looking time to the familiar member of the pair.

The effect found for the bilingual group replicates the novelty preference found at the same age with the fast Incidental Mutual Exclusivity task of Experiment 4. This convergence of results across different procedures makes us feel confident that bilinguals can perform such a task. In this particular experiment, the fact that did not show clear signs of extracting the label-category association in the Broad Categorization Trials makes the interpretation of the bilinguals result on IME trials problematic. However, it does not make it secondary. Results suggest that when the linguistic context is unambiguous, bilingual infants can understand that the learning task changes and concerns another category-label association, even while their learning process for the first association is still not completed. Thus, bilinguals seem to be flexible enough to use recently acquired information in order to disambiguate the referent of another label.

Another interesting finding is that the gaze pattern that bilinguals exhibited in the IME trials is compatible with the mechanism that Halberda (2006) described to be at the basis of the the Mutual Exclusivity strategy. He argued that infants might use a kind of reasoning, a disjunctive syllogism, to disambiguate the referent of a new label when one of the candidates is known. In his studies, adults and children first attended to the known object of the pair, and then they shifted to the novel image, suggesting a proper process of exclusion of an alternative. Perhaps in our task we are seeing exactly the same mental process while bilinguals are processing information which is not consistent with what they were familiarized with. In other words, our results suggest that bilinguals need not rely on the sedimented knowledge of a proper lexical item (as the one generally employed in 'classic' ME studies) in order to apply a process of double-checking. This speaks for a level of flexibility in bilinguals yet unproven.

In IME trials, monolinguals tended to look to the familiar picture after the label presentation. Again, this result is consistent with the familiarity preference found in the IME trials of Experiment 4, and therefore, we feel confident enough to draw firmer conclusions about this aspect of their behavior. What may account for it?

The data may suggest a persistent lack of flexibility in switching between tasks. Because in most of the task trials the label required them to attend to the familiar image, it might be difficult for them to shift from a predominant response to a novel task demand. This sort of perseveration is not necessarily a negative aspect for lexical acquisition. After all, lexicon is simply the crystallization of the relationship between labels and referents. Resistance to change and stability of mapping may be important ingredients for successful acquisition strategies. Perhaps the lack of flexibility may be beneficial for monolinguals to close in their lexicon faster than bilinguals, just as an increased flexibility in acquisition strategies may be useful for bilinguals to understand better what the learning task is about, at the expenses of a slower overall development of the vocabulary of each of the languages that they have to acquire.

Overall, Experiment 5 showed that when monolinguals are tested with a slow categorization task that requires certain stability, they can map new labels to categories. Bilinguals, instead, showed an advantage in those tasks that benefit from having more flexibility in switching from one strategy to another. In Experiment 6, we will try to explore the role of labelling on categorization further as well as the effects of linguistic experience on referential expectation. We will test infants with an ambiguous visual category that can be categorized either as a broad category or as two narrow categories. One label will be consistently assigned to each of the categories.

The experiment aims at exploring the causal role of labelling in category acquisition. We ask whether labels can direct infants' attention to features of the stimuli that so far where not relevant for categorization: the sub-categorical feature of leg numbers that infants failed to exploit in the Narrow Categorization Trials of Experiment 5. The experiment will also allow us to probe the role of linguistic experience in guiding referential expectations. Specifically, we will ask more directly whether bilinguals and monolinguals form the same referential expectations when two labels are presented.

3.3. Experiment 6: Induction of two simultaneous categories with a slow categorization task at 18 months of age

In Experiment 6, we test 18-month-old bilinguals and monolinguals in their ability to acquire two simultaneous categories in a familiarization/guided-looking procedure. Infants will be presented with two visual categories (2-legged and 4-legged dinosaurs) and two labels consistently applied to each of the category groups. We want to explore if labels can guide infants to categorize on the bases of a feature of the stimuli about which they showed no particular sensitivity in Experiment 5 (*vide* the results in the Narrow Categories Trials). Another focus of the experiment is to explore how linguistic background shapes infants' referential expectations. To that purpose, we will create a linguistic context clearly signaling than the two labels belong to the same language. Previous evidence coming from the individuation field (Xu, 1999) and from the categorization literature (Waxman & Braun, 2005; Plunkett et al., 2008) suggest that when more than one label is used, infants tend to infer that more than one referent is present. When a single label is used to refer to all category members, infants tend to form a single category (Waxman & Braun, 2005). The interest of the current experiment is to see whether when labels are meant to draw infants' attention to very subtle aspects of the definition of a category, they can indeed exploit language to refine their word acquisition strategies.

3.3.1. Materials and methods

3.3.1.1. Participants

Twenty-eight full term 18-month-old infants were retained for analysis: 17 Monolinguals (6 girls. Mean age: 18;07, Range: 17;12 – 18;30) and 11 Bilinguals (3 girls. Mean age: 18;14, Range 17;19 – 18;19). All participants were healthy and free from birth complications according to the parental report.

Eighteen additional infants (10 monolinguals and 8 bilinguals) were tested but not included in the analysis because they cried or refused to be seated (8), they were inattentive (2), there was parental interference (1), the equipment failed (4), the infant moved out of the area of eye capture of the Eye-Tracker (1), the infant had hearing problems (1), or the infant did not meet the data filtering criterion (see § 3.2.2.1.) (1).

The total rejection rate was 39%. Again, we stress the fact that many participants ran another unrelated task before being tested for this experiment. This fact may reduce their attention and hence underestimate the size of any potential effect.

The bilingualism definition and the recruitment procedure was as in Experiment 1.

3.3.1.2. Materials

3.3.1.2.1. Stimuli

We selected a subset of 20 images from Experiment 5: 8 were 2-legged dinosaurs, 8 were 4-legged dinosaurs, and 4 were animals from novel categories. Each pictures was presented only once in the experiment.

Additionally, we recorded a silent video of a smiling woman to be used in test trials. The actress was the same as in Experiment 5

The auditory stimuli and the attractor videos of the familiarization were the same as in Experiment 5.

3.3.1.2.2. Apparatus

The experimental set up was identical to the one described in Experiment 1.

3.3.1.3. Procedure

The procedure was identical to that of Experiment 5 except for the changes described below.

In the *familiarization phase*, infants were presented with eight pictures of the visual category dinosaurs that could be decomposed in two subsets of pictures recognizable by their differential features (4-legged dinosaurs and 2-legged dinosaurs). These pictures were presented in blocks of four. In each block, one label was consistently paired with the all the members of the subcategory kind, and another label was paired with the members of the other subcategory. The association between the labels and the subcategories was counterbalanced across participants (see Figure 61).

Otherwise, the structure of the familiarization trials and the timing were the same as in Experiment 5.

Condition	Auditory Stimuli	Label	Referent	Visual Stimuli
Familiarization	"Ostres! Això és un MAPU" x 3	Label 1	4 legs Dinosaur	
Familiarization	"Mira! un MAPU!" x3	Label 1	4 legs Dinosaur	
Familiarization	"Oh!Veus el MAPU?" x 3	Label 1	4 legs Dinosaur	
Familiarization	"Apa! És molt bonic aquest MAPU!" x 3	Label 1	4 legs Dinosaur	
Familiarization	"Ostres! Això és un DOTI" x 3	Label 2	2 legs Dinosaur	
Familiarization	"Mira! un DOTI!" x 3	Label 2	2 legs Dinosaur	
Familiarization	"Oh!Veus el DOTI?" x 3	Label 2	2 legs Dinosaur	
Familiarization	"Apa! És molt bonic aquest DOTI!" x 3	Label 2	2 legs Dinosaur	
Broad Category Conditon	Doti	Label 2	2 legs Dinosaur	
Narrow Categories Condition	Doti	Label 2	2 legs Dinosaur	
Switch Condition	Doti	Label 2	None / 4 legs Dinosaur	

Figure 61. Experiment 6 design. Infants were presented with 14 trials organized in 8 Familiarization Trials and 6 Test Trials (Two test trials per condition).

<u>The test phase</u> was identical to Experiment 5 in structure, but the kinds of test trials acquire a different role due to the change in familiarization (now with two labels consistently associated with two subcategories). As in Experiment 5, in the Broad Category Trials, infants saw a new dinosaur member (2 or 4-legged) and a picture of a novel category (E.g. a Fish). Infants listened to the label

that corresponded with the specific subclass of dinosaur currently shown, as induced in the familiarization trials. In the Narrow Categories Trials (2 trials), a new member of each of the familiar subcategories (a novel 4-legged dinosaur and a novel 2-legged dinosaur) were shown, but only one of them was labelled in each trial. The labeling was consistent with the familiarization. Therefore, the Narrow Category trials directly test whether infants tied the two familiarized labels with the two subcategories.

Finally, in place of the IME trials, we introduced two *Switch Trials.* In them, infants were presented with the same visual contrast as in a Broad Category Trial--that is, a novel dinosaur (E.g. 4-legged dinosaurs) and a picture of a novel category (E.g. a bird). However, the label that infants listened corresponded to the familiar category not on display (E.g. a 2-legged dinosaurs). The purpose of these trials is to test if, in case infants succeeded in the Broad Category Trials but failed in the Narrow Category trials, they treated both labels as synonyms for the same visual category (that is, dinosaurs). If this were the case, then they should be willing to accept that a label for a 2-legged dinosaur can be good for a 4-legged dinosaur, but not for a bird. (see Figure 61).

3.3.2. Results

3.3.2.1. Scoring

Data were obtained and analyzed as in Experiment 5.

In this experiment, in order to calculate the PLTT score, the target image was defined as the labelled picture of the trial. In Switch Trials, where the labelled referent is not present, we set that the target image was the familiar member of the pair.

For the time-course analysis of the Post-Label period in Switch trials we ran multiple two-tailed t-test comparisons instead of one-tailed t-tests because the two directions of the response were possible.

3.3.2.2. Results

Preliminary analyses on the Familiarization Trials did not detected any effect of Sex (Girls or Boys) ($P = 0.91$), Label (Doti or Mapu) ($P = 0.32$) or Visual Category (2-legged or 4-legged Dinosaurs) ($P = 0.55$) on infants' PLTT scores. Therefore, trials were collapsed across these conditions.

To correct for baseline preferences for one category, in test trials we compared infants' performance against their Pre-Label preference scores.

3.3.2.2.1. Familiarization Trials

The results reported here include the averaged responses to all the Familiarization trials collapsed.

We ran a two-way mixed ANOVA on the PLTT in Familiarization trials with Time Window (Pre-Label/Post-Label 1/Post-Label 2/ Post-Label 3) as a within–participants factor, Linguistic background (Bilinguals/Monolinguals) as a between–participants factor and

Participants as a random factor nested in Linguistic background. The results showed a main effect of Time Window (*M PLTT Pre-Label* = 66.72, *M PLTT Post-Label 1* = 75.53, *M PLTT Post-Label 2* = 63.11, *M PLTT Post-Label 3* = 63.08; $F(1,78) = 8.24$; $P < 0.0001$). Post hoc Scheffé tests revealed that infants increased the PLTT between the Pre-Label and the Post-Label Period 1($P < 0.03$). However, they significantly reduced the PLTT from the Post-Label 1 to the Post-Label 2 period ($P = 0.0008$).

There was no effect of linguistic background ($P = 0.57$), nor an interaction between Time Window and Linguistic background ($P = 0.52$). Thus, the results show that both bilinguals and monolinguals increased the PLTT during the period following the label presentations. No differences between groups were found.

Then, a time-course analysis on the PLTT of the familiarization trials was conducted. The analysis revealed some long sequences in which bilinguals and monolinguals increased the PLTT above chance. In monolinguals, it was found a long sequence of 7300 ms (Between 500 ms and 7700 ms after the trial onset and ending) was found to be significant. There was a second time window from 7900 ms to 11800 ms after the trial onset, although it was interrupted by some t-tests at $p <0.1$. For bilinguals, the analysis found three sequences. The first and longest sequence, started 500 ms after the trial onset and ended 7300 ms after the trial onset (that is, 6800 ms). The second sequence extended from 9100 ms to 11600 ms after the trial onset (that is, 2600 ms). The third sequence started around

12500 ms and ended 16100 ms after the trial onset (3700 ms). Finally, around the end of the trial, a sequence (600ms) was found in which the dynamics of the two groups differed (that is, between 13400 and 13900 ms after the trial onset). This effect do not fall inside any of the predefined 2500 ms time windows that follow a label presentation (see Figure 62).

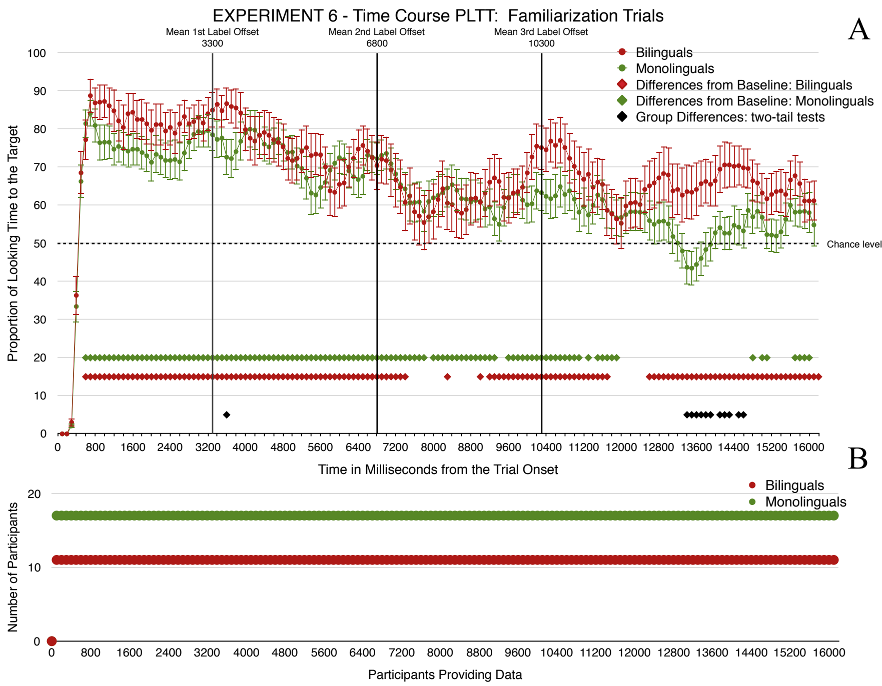


Figure 62. **A.** Symbols representing the PLTT across time in the Familiarization Trials of Experiment 6. Vertical lines represent the Mean offsets of the three label presentations. The Horizontal line indicates the Chance level. Also are shown symbols representing the sequences of significant one-tail t-test at p< 0.05 different from Chance level for the bilingual and monolingual groups and the significant sequences of two-tail t-test at p< 0.05 indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

The ANOVA and the time-course results together show that both groups attended to the familiarization trials similarly. The PLTT was higher after the first label presentation and it decreased across the trial.

3.3.2.2.2. Broad Category Trials

The results reported here include the averaged responses to the two Broad Category Trials collapsed. As in Experiment 5, we restricted all our analyses to the first label presentation.

For each participant, an averaged score for the Pre-Label and the Post-Label 1 period was obtained. Then, we ran a two-way mixed ANOVA on the PLTT in Broad Category Trials with Time Window (Pre-Label/Post-Label 1) as a within–participants factor, Linguistic background (Bilinguals/Monolinguals) as a between–participants factor and Participants as a random factor nested in Linguistic background. There were no main effect of Time Window ($P = 0.84$) or Linguistic Background ($P = 0.51$), nor an interaction ($P = 0.44$). (see Figure 63).

Then, a time-course analysis on the PLTT in Broad Category Trials was conducted to explore the dynamics of the response. Each time-bin in the 2.5 s of the Post-Label 1 time window of interest was compared against the Mean PLTT score for the Pre-Label period. An independent baseline score for the monolingual and the bilingual group was obtained.

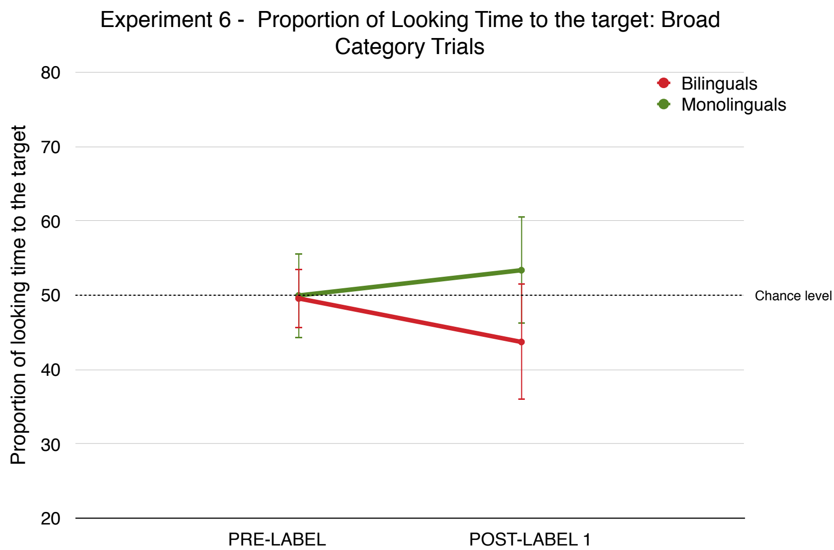


Figure 63. Mean PLTT during the Pre-Label and Post-Label period for the Broad Category Trials of Experiment 6. Bars represent SEs. * indicate effects at $p < 0.05$

The time-course analysis did not revealed any time window in which bilinguals or monolinguals increased their PLTT scores above the baseline. No group differences were found (see Figure 64).

Overall, the analysis of the Broad Category Trials suggest that infants were not able to identify the familiar members in the test pairs. There is no evidence that infants could map any label to the broad category Dinosaurs. Yet, the visual inspection of the temporal course again showed an initial strong preference for the familiar member of the pair (at least for bilinguals). It is possible that the selection of another baseline period where the gaze behavior was more stable might lead to different results.

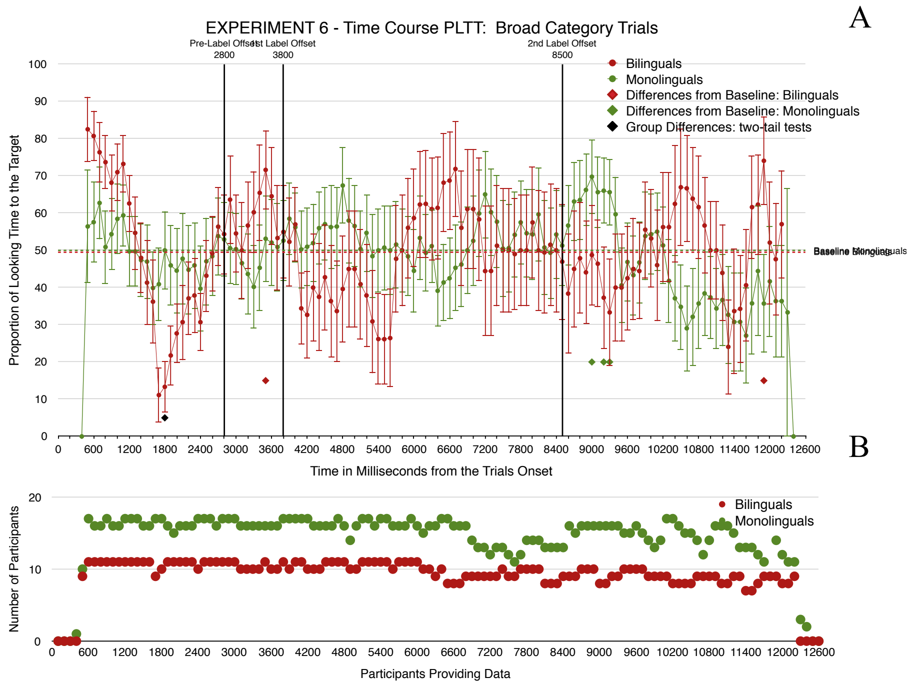


Figure 64. **A.** Symbols representing the PLTT across time in the Broad Category Trials of Experiment 6. Vertical lines represent the offsets of the two label presentations. Horizontal lines indicate the Pre-Label reference score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p< 0.05$ different from the Pre-Label score in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p< 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

3.3.2.2.3. Narrow categories Trials

The results reported here include the averaged responses to the two Narrow Categories Trials collapsed.

A two-way mixed ANOVA on the PLTT in Narrow Category Trials with Time Window (Pre-Label/Post-Label 1) as a within–participants factor, Linguistic background (Bilinguals/

Monolinguals) as a between–participants factor and Participants as a random factor nested in Linguistic background was computed. There was no effect of Time Window (P = 0.32) but there was an effect of Linguistic Background (M PLTT Bilinguals = 40.28, M PLTT Monolinguals = 54.91; F(1,24) = 5.45; P = 0.03). Post hoc Scheffé tests revealed that monolingual infants looked more to the target images than bilinguals (P = 0.03). This result might be driven by an incomplete randomization in the bilingual group. There was no interaction between Time Window and Linguistic Background (P = 0.58) (see Figure 65).

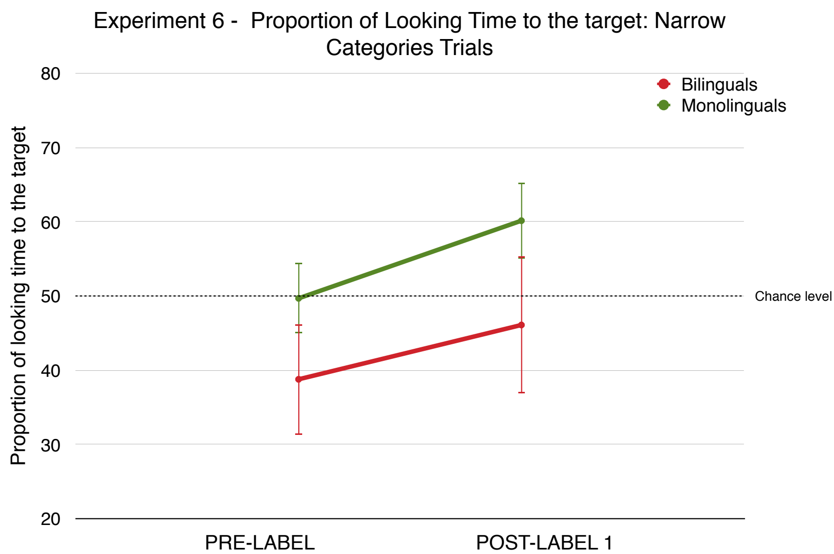


Figure 65. Mean PLTT during the Pre-Label and Post-Label period for the Narrow Categories Trials of Experiment 6. Bars represent SEs. * indicate effects at p <0.05

The time-course analysis revealed that monolinguals increased the PLTT above their baseline score after the first label presentation.

The time window started immediately after the first label presentation and was 1 s long (That is, from 3700 ms to 4600 ms after the trial onset). There was also a region in which the dynamics of bilinguals and monolinguals differed. The difference started at the next time-bin following the first label presentation and extended in time for 900 ms (From 3900 to 4700 ms after the trial onset). The visual inspection of the temporal course suggests that bilinguals might also present an increase of the PLTT (see Figure 66).

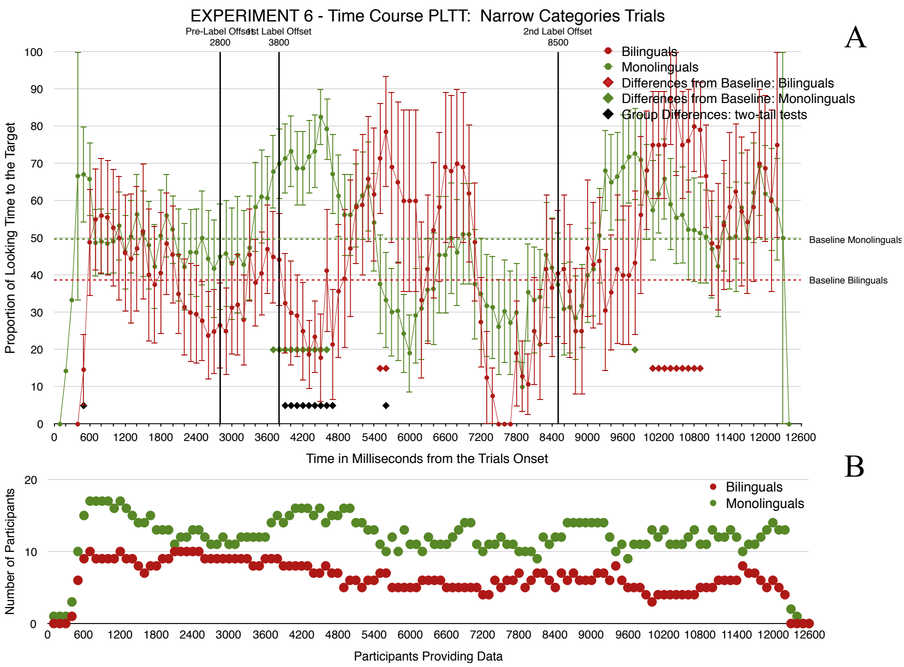


Figure 66. **A.** Symbols representing the PLTT across time in the Narrow Categories Trials of Experiment 6. Vertical lines represent the offsets of the two label presentations. Horizontal lines indicate the Pre-Label reference score for each group. Also are shown symbols representing the sequences of significant one-tail t-test at $p < 0.05$ different from the Pre-Label score in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Still, we need to be cautious about this interpretation due to the small sample size. This result suggest that monolinguals successfully learned the two label-subcategory pairings.

3.3.2.2.3. Switch Trials

The results reported here include the averaged responses to the two Switch Trials collapsed.

A two-way mixed ANOVA was computed on the PLTT in Narrow Categories Trials with Time Window (Pre-Label/Post-Label 1) as a within–participants factor, Linguistic background (Bilinguals/ Monolinguals) as a between–participants factor and Participants as a random factor nested in Linguistic background. There were no effects of Time Window (P = 0.23) or Linguistic Background (P = 0.31), nor an interaction (P = 0.95) (see Figure 67).

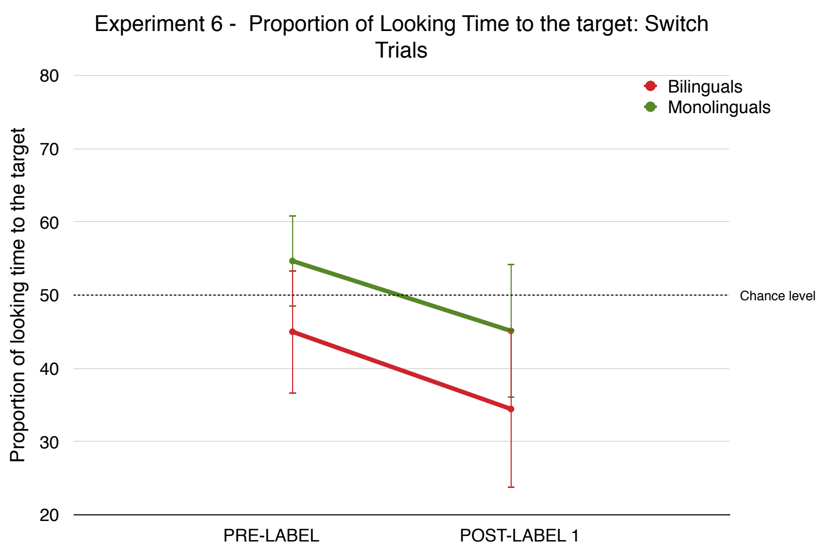


Figure 67. Mean PLTT during the Pre-Label and Post-Label period for the Switch Trials of Experiment 6. Bars represent SEs. * indicate effects at p <0.05

In the time-course analysis, because both images were potential referents of the label, we decided to run two-tailed t-tests for the comparison against the Pre-Label reference score. We found no region in which either group gave signs of successfully mapping the label to any of the images (see Figure 68).

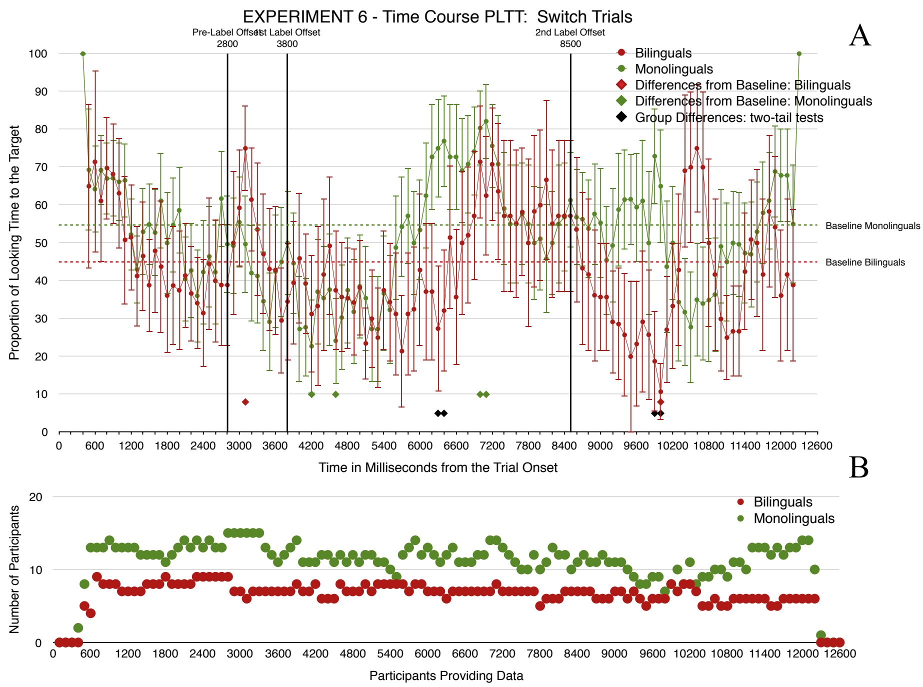


Figure 68. **A.** Symbols representing the PLTT across time in the Switch Trials of Experiment 6. Vertical lines represent the offsets of the two label presentations. Horizontal lines indicate the Pre-Label reference score for each group. Also are shown symbols representing the sequences of significant two-tail t-test at $p < 0.05$ different from the Pre-Label score in monolinguals and bilinguals and the significant sequences of two-tail t-test at $p < 0.05$ indicating group differences. Bars represent SEs. **B.** Plot representing the number of participants providing data in each time-bin.

Together, the results of the Switch trials suggest that infants did not have any systematic referential expectation when the labels of the two narrow categories were switched. Results might suggest that infants were not particularly surprised by the label-subcategory violation.

3.3.3. Discussion

Note that the data we discussed in Experiment 6 are preliminary results and thus, we need to be cautious when interpreting them.

In this experiment, we were interested in exploring the role of labelling in guiding infants' categorization, as well as its interplay with linguistic experience.

We found that when two labels were consistently paired with two subcategories, monolinguals tended to extract two categories. This result replicates the findings of (Plunkett et al., 2008) on 10-month-old monolinguals using a similar procedure.

We speculate that these results suggest a certain causal role of labels on the acquisition of two categories. In the Narrow Category Trials of Experiment 5, infants were also presented with a 2-legged vs a 4-legged dinosaur pair, monolinguals did not show to be sensitivity to the feature “number of legs”. Thus, it is possible that in Experiment 6 labels guided infants’ attention towards the relevant features of each subcategory.

Still, why monolinguals did not show signs of learning in the Broad Categorization trials remains to be explained. If infants were learning two simultaneous label-subcategory pairings, theoretically we should expect that they succeed in the Broad Category Trials as well.

In this experiment, results did not show that bilinguals learned in any of the test conditions, although no significant differences between bilinguals and monolinguals were found. Moreover, in the broad category trials, bilinguals showed a strong initial familiarity preference followed by a novelty preference. It might be possible that the selection of another baseline period that excluded this initial dynamics may lead to different conclusions.

A final methodological remark is that, in Experiment 6, we decided to introduce a short video that preceded the stimuli presentation on each trial, as compared to Experiment 5 in which it was only presented in two out of the eight familiarization trials. We consider the possibility that this modification may have increased the memory load with respect Experiment 5. It is not impossible that a long distance between presentations made the encoding and retention of information especially difficult.

4. GENERAL DISCUSSION

The aim of this dissertation was two-fold. First, we were interested in exploring infants' ability to learn names for categories as well as their flexibility to exploit recently acquired information and use it online to disambiguate the meaning of other labels.

Second, we were interested in understanding the influence of linguistic experience on word learning strategies. Here, we understood linguistic experience as the knowledge about the language/s that the infant is exposed to and the potential byproduct it may have on cognitive abilities such as executive functions.

In Experiments 1-4, we tried to understand whether the ability to rapidly map labels to objects on the basis of a few presentations of the contingency, recently discovered in 8-month-old monolinguals (Saksida, 2014), could be extended to the task of learning labels for categories with the same immediacy and efficacy.

To address this question, we tested monolingual and bilingual infants at 8, 15, and 19 months of age with an Infant Contingent Categorization Task and an Incidental Mutual Exclusivity task. In the former task, infants were familiarized with a few brief presentations of pairs of images that belonged to distinct categories. In each presentation, a member of each category shown. One of the categories was consistently paired to a novel label across trials. That group becomes the target category. Orientations towards the target category after the label presentation triggered a visual reward. The members of the other category were not paired to any label and could not trigger the reward. Thus, in test trials we expect that those infants who learned the label-category pairing will tend to look towards the target category when they listen to the same label than in previous trials. This task tests infants' ability to map new labels to categories. The Incidental Mutual Exclusivity task was a complementary task added to the Infant Contingent Categorization Task. There, a novel member of each category were shown as in previous trials but this time infants listened to a novel label. This task tested if infants could map the novel label with the category that had not been associated with any label in previous presentations, thus showing ability to disambiguate the referent of a new label.

Bilinguals and monolinguals differ in some essential aspects of their linguistic experience and their linguistic input (see § 1.2.). For all these differences, an adaptive word learning heuristic in one context may not fill the requirements of the other environment. The comparison between bilinguals and monolinguals allowed us to assess the role of the linguistic environment at each age to solve these tasks. The comparison across ages gave us the chance to assess the role of general maturation of cognitive abilities (presumably similar in monolinguals and bilinguals) as well as the importance of linguistic exposure (clearly different between bilinguals and monolinguals) for how category learning changes in infants, whether as a direct result of the acquisition of linguistic knowledge or as an indirect consequence of the modification of cognitive abilities induced by language.

In consideration of the results obtained in Experiments 1-4, we felt the need to test 18-19 month-old infants with another procedure. Thus, in Experiments 5-6, we tested monolinguals and bilinguals of this age with a slow categorization task, allegedly more akin to word-learning.

Because the age of bilingual and monolingual participants was balanced within each experiment, we expect that any difference that emerged between groups is directly or indirectly attributable to linguistic environment.

In the next sections we will try to explain how the factors we controlled in our experiments, or an interplay among them, can account for our findings.

2.6.1. Categorization abilities

We implemented two different procedures aimed at testing categorization abilities.

In a set of experiments (Experiments 1-4), we used a “fast” categorization task that proposed rapid repeated presentations of the relationship between a label and a category.

In Experiment 1, we demonstrated that our adaptation of a label-object mapping task was useful to induce label-category mapping at 8 months of age. This result extends the rapid learning of object-label pairs found by Saksida (2014) to the acquisition of labels for categories. This is no trivial result, because infants need to construct their knowledge of the language-world interaction, not only by acquiring the names for objects, but especially by acquiring ways to understand the relationship between concepts and their extensions. Words for categories are the mediators for this relationship and an essential mean for the expression of thoughts. As we argued before, the appreciation of label-category links is an important milestone towards understanding words as full lexical items (see Preface).

In this experiment, both bilinguals and monolinguals succeeded at learning a new label-category pairing. Our procedure, which presents test trials interspersed with label-category learning trials, allowed us to also try to understand the dynamics of learning. To do that, we analyzed both when learning occurred *during the time-frame of the experiment* and when learning occurred *within the time-frame of each trial.* Inside the experiment, we found that learning already occurred after few presentations of the label-category relation. The fact that signs of learning occurred in the first part of the experiment suggests that the pairing was a fast process. The questions of how stable this learned relationship is (that is, how deeply in memory it is retained), as well as the specific linguistic nature of this learning (that is, how much is this learning is *really* word learning, as opposed to being the acquisition of a label-category association), remain open.

In Experiment 2, we introduced two modifications. First, we shortened the experiment length to understand how much we could push infants' fast-mapping abilities. Second, we added an adaptation of a Mutual Exclusivity task—a novel task, which we called *Incidental Mutual Exclusivity,* aimed at assessing how much infants could flexibly exploit a label-category pairing during the learning process, in order to acquire novel knowledge. The results of Experiment 2 replicated the findings of Experiment 1: infants could learn to map a novel label to a category after a few brief presentations. However, we also found that that bilinguals outperformed their monolinguals peers in the fast categorization task in both speed of mapping and strength of the response. Bilinguals showed signs of learning the label-category contingency already during the first half of the experiment and they also presented longer periods in which they looked at the target.

The bilingual advantage in the acquisition of labels for categories at 8 months is congruent with evidence showing that bilingual infants outperform their monolinguals peers in tasks that require executive functions at the preverbal stage (Kovács & Mehler, 2009a; Kovács & Mehler, 2009b). As we argued in (see § 2.1.), our categorization task is more dynamic than the "standard" familiarization/ preferential looking tasks. In the Infant Contingent Categorization task, the short presentations and the presence of test trials interspersed with the learning trials pose a more demanding challenge to infants, likely to engage their executive functions more fully. Thus, it is possible that in our task, bilinguals may benefit from more mature executive abilities and may adapt their learning strategy to its demands in a way that monolinguals may not. For this reason, they may acquire the label-category contingency faster. This is a clear effect of the linguistic environment on general learning abilities, and may unveil a novel cognitive consequence of bilingualism in word learning, so far not explored.

To explore if higher cognitive resources and more linguistic experience would improve label-category mapping, in Experiment 3 we tested 15-month-olds with the same Infant Contingent Categorization task. This experiment gave more paradoxical findings. We found that, despite the good performance at this task at younger ages, 15-month-old bilinguals and monolingual infants did not give clear signs of acquiring the labels for categories, nor did they react to the IME task. Several reasons may account for the failure at this age. It is possible that this procedure is not suitable for testing older infants. It is also possible that older infants were doing something *better* than their younger peers and that, for this reason, they were *worse* in our tasks. It has to be noted that for each of four trials of familiarization—which presented the category image with the reinforcement—there were two un-reinforcement test trials. Possibly, 15-month-olds were tracking the consistency of the label-category pairings better than their 8-month-old peers, monitoring the cross-situational occurrence of labels and rewards across trials (Smith & Yu, 2008). If this was the case, the test trials interspersed with the familiarization trials, which by definition were not reinforced, may have disrupted their learning. If this were the case, then testing older infants, who track the probability of occurrence of even distant events better (Marchetto & Bonatti, 2013), should also fail in our task. A final speculation is that at 15-months, infants may start to understand the Infant Contingent Categorization Task as a true word learning task, as opposed to simply looking at the relationship between a sound label and a set of images. This new interpretation of the task may make it more conservative as to what kind of evidence is sufficient to establish that a sound label is a proper word. This age may be the transition between looking at these tasks as contingent learning tasks and looking at them as proper lexical tasks, with the added processing loads that this passage likely involves. The behavior of 19-month-old bilingual and monolingual infants in our fast categorization task may shed some light on the failure at 15 months. This, besides the intrinsic interest of following the developmental trajectory of fast category learning, was the motivation for Experiment 4. Some interesting results emerged.

First, we found that, again, 19-month-old bilinguals could extract the label-category pairing. Signs of learning were present also in monolinguals, although to a lesser extent. This in itself is an unusual result, considering the literature, which would lead us to expect the opposite pattern, if any. We will come back to this point later. What we want to focus on now is the developmental trajectory which emerges when the acquisition of the label-category relations is compared across ages. Such a trajectory seems to follow a U-

Shaped curve, with an early success, a period of failure, and a second, later period of success in the task, although with a different learning profile, which we tried to describe in our analyses and comment about below.

Second, again, bilinguals showed an advantage for fast categorization compared to monolinguals. This aspect of our data strengthens the possibility that the successful performance in our task may require a certain degree of flexibility, in order to rapidly attend to the label and exhibit a response within the trial time.

Third, at 19 months the acquisition of the label-category relationship in bilinguals emerged in the second half of the experiment. By contrast, younger bilingual infants had already shown signs of learning during the first half of the experiment. We hypothesize that this result might suggest that 19-month-old infants approach the fast categorization task in a different way, perhaps charging labels with richer linguistic content than at previous ages. This may make the task more demanding, requiring deeper processing resources which translated into a slower acquisition of the same label-category relations.

The meaning of this result needs to be explored further. For the time being, we want to focus on the fact that, by itself, it already shows that the failure at 15 months is not necessarily due to a difficulty in learning sparse label-category relations interspersed with non-reinforced trials. Something else must be happening at 15 months, accounting for the U-shaped developmental profile of the success in our task. Further research is needed to clarify the exact reasons for the failure at 15 months. Yet, we suggest that current pattern of results could indeed be driven by a developmental shift from understanding labels signaling commonalities between members of a category, to understanding labels as words with a fully developed lexical meaning. We want to pause and discuss this possibility further.

In the process of identifying the relationship between labels and referents, phonological knowledge plays an important role. Indeed, we know that phonological knowledge may have an effect on word learning. For example, Werker, Cohen, Lloyd, and Stager (1998) showed that 14-month-old monolinguals could learn two label-object associations when the labels did not sound similar. However, when infants of the same age were tested with similar sounding words, they failed to learn the associations (Stager & Werker, 1997). Interestingly, the failure to create the links between labels and objects was found to be independent from the ability to discriminate these similar-sounding labels (Stager & Werker, 1997). It is not until 17 months of age that monolinguals could learn to map two similar sounding labels to two objects (Werker, Fennell, Corcoran, & Stager, 2002). This ability seems to be delayed in bilinguals, who do not succeed in that task until the age of 20 months (Fennell et al., 2007). It is possible that, at an age in which infants begin the transition from being good discriminators of verbal sounds of any language to focusing on sounds and words of their own target language, the processing of the images, the novel labels, and the establishment of links might be excessive for non proficient word learners such as 15-month-olds (Stager & Werker, 1997). At this age, infants may begin treating our tasks as involving the acquisition of lexical items for a valid language and they may have trouble identifying what constitutes a "valid" word form in it.

Going back to 19-month-olds, the last result we found related to categorization is that the percentage of exposure to a second language predicts whether infants link labels with categories in our Infant Contingent Categorization Task. After listening to the label, the bigger the exposure to a second language was, the stronger infants' tendency to look at objects of the target category. This relation was not observed at earlier ages. Although we do not know which specific characteristics of the bilingual input guided this result, we speculate that infants with a more balanced exposure to two languages may positively exploit the stronger interference between them. Because they need to monitor the two languages in a more accurate way, those bilinguals may develop stronger executive control abilities compared to infants exposed to less or no linguistic conflict. This result is another sign that linguistic environment moulds some aspects of the acquisition of the label-referent relationship. It may also be further confirmation that 19-month-olds, like 15-month-olds, interpret the task as a proper linguistic task. The difference might be that 19-months-old infants perhaps have a stronger grip on the prerequisite representations (lexical or phonological) that can allow one to exploit linguistic experience.

Experiments 5-6, using a different technique, tried to address the issue of how linguistic knowledge and linguistic experience influence the ability to learn names for categories. With that purpose, we tested 18 and 19-month-old bilingual and monolingual participants in a “slow” familiarization task more akin to word-learning and less demanding in flexibility. This procedure is meant to give a point of reference more similar to the current literature about labelling and categorization in infants. (e.g. Fulkerson & Waxman, 2007; Ferry et al., 2010; Ferry, Hespos, & Waxman, 2013).

By the time of the submission of this work, we lack of a sufficient group of bilingual infants. Therefore, these results must be considered preliminary. However, although we need to be cautious about their interpretation, these data gave us some clues to better understand our previous results.

In Experiment 5, we familiarized infants with a label-category pairing. Interestingly, in this experiment with less executive function demands we found that when 19-month-old monolinguals were tested, they could learn the label-category pairing. However, this time it was the bilingual group who did not show clear signs of learning.

In Experiment 6, we also found that labels seemed to have a role in guiding infants' attention. We found that when infants were presented with two labels that correlated with attributes of the category, monolingual infants could categorize on the basis of a contrast to which they had not shown to be sensitive in Experiment 5 (number of legs). This finding suggests that labels may have a causal role in grouping objects, perhaps strengthening hypotheses about the role of words in identifying objects (Xu, 2002), but at later ages and applied to category. Perhaps, what this result amounts to, though, is the fact that labels may help chunking groups of objects in memory, rather than being constitutive of the identification of the categories. Similar results were found by (Feigenson & Halberda, 2008).

Altogether, we saw that if we test 19-month-old infants with a task that requires some sort of stability of behavior, as might be the case of the slow categorization task, monolinguals perform equal to or better than bilinguals. Instead, we saw that when a task requires flexibility, bilinguals may show an advantage. Therefore, it is not the case that bilingualism per se favors category acquisition. Rather, bilinguals may be favored in those processes of word/label learning that require a more flexible response. We now turn to this aspect of our dissertation.

2.6.2. Flexibility in the acquisition of category labels

In Experiments 2 through 6, we also explored bilingual and monolingual infants' flexibility in using recently learned Label-Category pairing "online" to disambiguate the referent of another label in an Incidental Mutual Exclusivity task (Experiment 2-5) or in a switch task (Experiment 6).

In Experiments 2 and 3, 8- and 15-month-old bilingual and monolingual infants were presented with an ambiguous referential situation. In it, infants needed to exploit the relationship between label and category induced in previous trials in order to discover the referent for a novel label. Thus, our disambiguation task is built on the successful acquisition of a label-category relationship, together with infants' ability to exploit this new information right after (or during) its acquisition. In this sense, our task is quite different from standard versions of the Mutual Exclusivity task so pervasive in the literature, in which responses are based on the exploitation of stable, previously acquired knowledge of real lexical entries. Our task requires strong flexibility, so as to change the course of the learning strategy. Infants needed to rapidly process the acoustic input, reason during the brief trial time about the possible referent and shift from the predominant response in previous trials (look to the target) to a new one (look to the non-target). Recently, it has been shown that 4-month-old infants can disambiguate the referent of a new label with a similar procedure (Saksida, 2014).

Overall, results of Experiments 2-3 suggest that infants were not very systematic in their responses, although in some circumstances at 8 and 15 months we found signs of success. For example, in Experiment 2, we found that during the second half of the task, monolinguals decreased the looking time to the member of the category associated with the previously reinforced label (our measure of success). In Experiment 3 and 4, we also found partial success in bilinguals when we split by Target Category.

Our results were weaker than previous findings interpreted as the presence of ME reasoning at very young ages (Saksida, 2014).

In Experiment 4, we tested 19-month-old bilingual and monolingual participants with the same IME task. Interestingly, but unexpectedly, we found that bilinguals showed strong signs of orientation towards the category of the novel label following the label presentation. This finding suggests that bilingual infants could exploit the label-category pairing learned during the familiarization trials to bootstrap the learning of a second label-category pairing.

To our knowledge this is one of the first evidences that bilingual infants of this age can use an analog to the ME strategy to disambiguate the meaning of new labels within language online. Other studies on ME suggest that bilinguals tend not to apply this strategy to learning novel words (Byers−Heinlein & Werker, 2009; Houston−Price et al., 2010). We suggest that the difference in results lies in the differences in the tasks. In a “standard” ME task, infants are tested on their ability to use information that they already knew in order to disambiguate the referent of a novel label. Notice that in our task, infants have to learn and apply this new knowledge *during* the task. That may add a stronger executive component to our task. It is possible that flexibility in incidental learning, rather than the presence of stable lexical representations, is what our task really tests. It has been shown that bilinguals at different ages are better than monolinguals at switching between different tasks or rules (Bialystok, 1999; Bialystok & Martin, 2004; Kovács & Mehler, 2009a). Bilingual adults also suffer less switching costs when alternating between different rules (Costa, Hernández, & Sebastián-Gallés, 2008). It is possible that monolinguals cannot disengage from one kind of trial sufficiently quickly to reorient towards the novel target. Thus, a bilingual advantage in executive functions may facilitate the IME task.

An apparently counterintuitive finding was that in Experiment ## 19-month-old monolinguals did not seem to be able to solve our IME task. This is an age at which the literature attests that they use ME (Halberda, 2003; Houston−Price et al., 2010; Byers−Heinlein & Werker, 2009). In IME, we found the opposite pattern; in some cases monolinguals even increased their direction towards the familiar category induced in the familiarization trials, even when they had just listened to a novel label. However, this result is only apparently contradictory in the context of ME literature. If we pursue the hypothesis that the IME task requires flexibility for changing the course of learning, the inability to solve IME in 19-month-old monolinguals may simply show that in a fast situation such as that presented by IME they have difficulty inhibiting a predominant response.

To try to isolate the role of linguistic experience in the use of the IME, we also tested 19-month-old infants in a “slow” IME task.

Experiment 5 revealed that, again, bilinguals gave indications of applying IME reasoning, but not monolinguals, who nevertheless did learn the category induced by the familiarization. The result replicates the bilingual success at 19 months in Experiment 4, with a “fast” IME task. Furthermore, the gaze pattern of bilinguals in Experiment 5 was particularly revealing. Previous to the orientation towards the novel image, bilinguals showed a strong increase of looking time to the familiar image. This pattern is strongly reminiscent of a similar "check and discard" strategy revealed by Halberda (2003) in older children. The computations that allow infants to solve the IME task remain unclear, and go beyond the aim of the current dissertation. Still, we speculate that the underlying mechanism that may account for that behavior may well be the same: a kind of reasoning via disjunctive syllogism to discard an old referent and orient towards a novel referent (Halberda, 2003; Halberda, 2006), but this time applied to the task of learning words for categories. Thus, our bilingual participants may have used a logic domain-general strategy to disambiguate the referent of a new a label. Again, our findings highlight the complexity of even apparently simple learning tasks, and the difficulty of pinpointing a strategy that is unequivocally dependent on specific strategies for lexical acquisition. When inspecting how infants may bootstrap the acquisition of words for novel categories in a dynamical learning environment, we cannot avoid considering how non-linguist abilities, such as the flexibility allowed by well functioning executive functions, or some kind of proto-reasoning, may foster this process. At the same time, and probably for the same reasons, we cannot avoid considering linguistic background as a factor in any successful task of categorization.

5. CONCLUSIONS, LIMITATIONS AND FUTURE DIRECTIONS

Based on the results obtained in this dissertation, several conclusions can be reached.

First, by implementing a new procedure aimed at inducing rapid label-category mappings starting from 8 months, we showed that infants can map new labels to categories after a few brief presentations of the label-referent association.

Second, we found that 8 and 19-month-old bilingual infants gave better signs of learning such associations than monolinguals, when they were presented with the same fast categorization task. We argued that this result might be driven by an advantage in executive functions.

Third, we found that the developmental trajectory of the label-referent association for categories seems to undergo a shift between ages 15-19 months. We made the hypothesis is that this U-shape marks the transition from treating linguistic signals as labels to completing their status as potential words for a developing lexicon.

Fourth, 19-month-old bilingual infants were able to exploit a recently learned label-category pairing to bootstrap the learning of a novel label, regardless of the details of the category learning task they were given.

Fifth, older bilinguals may use a domain-general mechanism, perhaps akin to logical reasoning, such as disjunctive syllogism, to disambiguate the referent of new words for categories.

Altogether, the differences found between bilinguals and monolinguals suggest a more complex interplay between linguistic background, task demands, and linguistic knowledge on word learning strategies than currently acknowledged.

The current work presents some novel and relevant contributions. It also comes with many limitations that make us cautious in interpreting the results. In this section, we will discuss these limitations, their implications, and some potential approaches to overcome them in future work.

First, we argued that labels may have guided category acquisition by directing infants' attention towards category commonalities (Experiments 1-4) or by changing the attention to the dimensions that infants considered to make the category (Experiment 5 and 6). However, our experimental design cannot rule out the possibility that infants categorized on the basis of other features of the stimuli and of the procedure. For example, they may track visual features of the stimuli and make a perceptual categorization. In our opinion, the possibility exists but is remote. Infants waited until the end of the label before orienting towards the target category, suggesting that they did process the label. Furthermore, in some experiments we found a distinct pattern of responses in IME trials with respect to that of Categorization trials (E.g. Experiment 4). Also, it would be difficult to explain the failure of 15-month-olds in categorization, if this were based on simple visual features of the stimuli. Nevertheless, a direct control experiment to discard that possibility would be more convincing. One option would be to test infants with the same procedure, but without any linguistic labels, in silence or pairing the images with a non-linguistic stimulus. Another option would be to design a task in which the processing of the label, or lack thereof, leads infants to distinct referential expectations.

Another limitation concerns a set of assumptions about the minimal amount of information that infants need to solve our tasks. While we tried to design our tasks on the basis of existing procedures, we clearly had to make assumptions about the number of trials needed to acquire the labels, which may be erroneous. This especially is true for Experiments 5 and 6. In particular, in Experiment 6 we tried to induce two labels for subcategories, on the basis of four presentations of the relationship. The null results we obtained may be due to this factor. In general, a better inspection of this part of our procedure would be needed.

A third limitation was given by our stimuli. Unfortunately, as it turned out, most infants had a very strong preference for one category over another. Although baseline correction can reduce the effect of previous preferences over the results, it certainly does not eliminate it. Simply put, when infants so clearly prefer one category, it is difficult, if not impossible, to find a further increase in looking time towards that category after a label presentation. For this reason, we were obliged to run separate analyses in several experiments, which the clean identification of learning effects difficult. Ideally, more balanced stimulus material could have helped us in our inquiry.

Another point worth noting in our work as well as in many other studies present in the literature, is the absence of established criteria for selecting a reasoned baseline on the basis of which to compare the effect of labels on infants' oculomotor behavior. While a general problem, this is a particularly severe limitation in our studies, because we found that some of the most interesting parts of infants' behavior can be seen only at the micro-level of gaze dynamics across each trial, as opposed to general learning effects. In our analyses, we opted for a conservative stand, assuming that the full period before label onset had to be considered to compute a baseline. However, there are clearly other options which may have given different results. In some experiments (Experiments 5-6), we found strong dynamics during the baseline period, over which we averaged, that may have compromised a more accurate analysis. Although see (Waxman, Lidz, Braun, & Lavin, 2009 for a different view).

Most importantly, the most serious limitation of our studies depends on our inability to find sufficient bilingual samples to get to firmer conclusions in our experiments. We are aware of this issue, but we cannot do anything else but report what we found and suggest lines of further completion and investigation of the phenomena we tried to uncover.

Despite all these limitations, we believe that our work has brought to light a series of very interesting phenomena, which we hope will offer elements to better understand the interplay between linguistic experience and the complex learning problems that infants are faced with every day while learning the vocabulary that, in few years, will allow them to express the rich palette of thoughts that characterizes human beings.

References

Albareda−Castellot, B., Pons, F., & Sebastian-Galles, N. (2011). The acquisition of phonetic categories in bilingual infants: new data from an anticipatory eye movement paradigm. *Developmental Science*, *14*(2), 395-401.

Baillargeon, R., & Wang, S. H. (2002). Event categorization in infancy. *Trends Cognitive Sciences*, *6*(2), 85-93.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants. *Journal of experimental child psychology*, *64*, 3-26.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of child language*, *6*(2), 183-200.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3253-3258.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child development*, *70*(3), 636-644.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging*, *19*(2), 290-303.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science*, *7*(3), 325-339.

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan, the rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*, 103-119.

Bomba, P. C. (1984). The development of orientation categories between 2 and 4 months of age. *Journal of Experimental Child Psychology*, *37*(3), 609-636.

Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 115-129.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of experimental child Psychology*, *112*, 417-436.

Bosch, L., & Sebastian-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*, 29-49.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and speech*, *46*(2-3), 217-243.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, *21*(3), 343-348.

Byers−Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: infants' language experience influences the development of a word-learning heuristic. *Developmental science*, *12*(5), 815-823.

Carey, S. (1978). The Child as Word Learner. In M. Halle, J. Bresnan, & A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Standford child language Conference*, *15*, 17-29.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*(2), 282-298.

Clark, E. V. The principle of contrast: A constrain on language acquisition. In MacWhitney (Ed.), *Mechanisms of language acquisition* (pp. 1-33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, E. V. (2016). Two pragmatic principles in language use. In D. Matthews (Ed.), *Pragmatic development in first language acquisition*. Amsterdam: John Benjamins.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*(1), 59-86.

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in cognitive sciences*, *10*(3), 131-138.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *16*(1).

Davidson, D., & Tell, D. (2005). Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects. *Journal of experimental child psychology*, *92*, 24-45.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe functioning*. New York: Oxford University Press.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254-260.

EuropeanCommission. (2012). Europeans and their languages - Special Eurobarometer 386 / Wave EB77.1.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(29), 9926-9930.

Fennell, C. T., Byers−Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: the case of bilingual infants. *Child development*, *78*(5), 1510-1525.

Ferguson, B., Graf, E., & Waxman, S. R. (2014). Infants use known verbs to learn novel nouns: Evidence from 15- and 19-months-olds. *Cognition*, *131*, 139-146.

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*, 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 of the United States of America*, *110*, 15231-11535.

Frank, I., & Poulin-Dubois, D. (2002). Young monolingual and bilingual children's responses to violation of the mutual exclusivity principle. *International Journal of Bilingualism*, *6*, 125-146.

Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling, sounds and source of auditory input on 9-and 15-months olds' object categorization. *Infancy*, *4*(3), 349-369.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, *105*, 218-228.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual review of psychology*, *60*, 115-140.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*(2-3), 263-297.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual review of psychology*, *61*, 191-218.

Pi, G. (2015). The role of experience in perceptual reorganization: The case of bilingual infants. Universitat Pompeu Fabra, Barcelona.

Golinkoff, R. M., & Alioto, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition. *Journal of Child Language*, *22*(3), 703-726.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., Leslie, M., & Wenger, N. L. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, *28*(1), 99-108.

Goyet, L., Nishibayashi, L. L., & Nazzi, T. (2013). Early syllabic segmentation of fluent speech by infants acquiring French. *PloS one*, *8*(11), e79646.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, *1*(2), 67-81.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28*(2), 240-244.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23-B34.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive psychology*, *53*(4), 310-344.

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science*, *146*(3643), 549-551.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of child Language*, *39*(1), 1-27.

Houston−Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, *15*(2), 125-150.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research*, *5*, 69-95.

Kandhadai, P., Hall, D. G., & Werker, J. F. (2016). Second label learning in bilingual and monolingual infants. *Developmental science*.

Konishi, H., Pruden, S. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Categorization of dynamic realistic motion events: Infants form categories of path before manner. *Journal of Experimental Child Psychology*, *152*, 54-70.

Kovács, A. M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America* , *106*(16), 6556-6560.

Kovács, A. M., & Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*(5940), 611-612.

Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive psychology*, *67*(3), 130-150.

Mareschal, D., Powell, D., Westermann, G., & Volein, A. (2005). Evidence of rapid correlation−based perceptual category learning by 4−month−olds. *Infant and Child Development*, *14*(5), 445-457.

Markman, E. M. (1991). The whole object, taxonomic, and mutual exclusivity assumptions as ubitial constrains in word meaning. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thoought: Interrelations in development*. Cambridge, UK: Cambridge University Press.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121-157.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive psychology*, *47*, 241-275.

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive science*, *36*(7), 1157-1177.

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, *6*(2), 203-229.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143-178.

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. *Child development*, *65*(6), 1646-1662.

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental science*, *10*(6), 719-726.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *The Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756-766.

Cesana-Arlotti, N. (2015). *The origin of logical concepts.* Universitat Pompeu Fabra, Barcelona.

Pearson, B. Z., & Fernandez, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language learning*, *44*(4), 617-653.

Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language learning*, *43*(1), 93-120.

Petitto, L. A., Katerelos, M., Levy, B. G., Gauna, K., Tétreault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: implications for the mechanisms underlying early bilingual language acquisition. *Journal Child Language*, *28*(2), 453-496.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665-681.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastian-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10598-10602.

Quine, W. V. O. (1960). *Word and objects*. Cambridge, UK: University Press.

Quinn, P. C., & Eimas, P. D. (1996). Young infants' use of facial information in the categorical differentiation of natural animal species: The effect of inversion. *Infant Behavior and Development*, *19*(3), 381-384.

Quinn, P. C., Slater, A. M., Brown, E., & Hayes, R. A. (2001). Developmental change in form categorization in early infancy. *British Journal of Developmental Psychology*, *19*(2), 207-218.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-innanimate distinction. *Psychological bulletin*, *127*(2), 209-228.

Ramon-Casas, M., Swingley, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, *59*(1), 96-121.

Roberts, K., & Jacob, M. (1991). Linguistic versus attentional influences of nonlinguistic categorization in 15-months-old infants. *Cognitive Development*, *6*, 355-375.

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*, 95.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child development*, *75*(5), 1387-1401.

Robinson, C. W., & Sloutsky, V. M. (2007). Linguistic labels and categorization in infancy: Do labels facilitate or hinder. *Infancy*, *11*(3), 233-253.

Romaine, S. (1995). *Bilingualism*. Oxford: Wiley-Blackwell.

Rostad, K., Yott, J., & Poulin-Dubois, D. (2012). Development of categorization in infancy: Advancing forward to the animate/ inanimate level. *Infant Behavior and Development*, *35*(3), 584-595.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Saksida, A. (2014). *Word learning in the first year of life.* Scuola Internazionale Superiore di Studi Avanzati, Trieste.

Sebastian-Gallés, N. (2010). Bilingual language acquisition: where does the difference lie. *Human Development*, *53*(5), 245-255.

Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: role of early exposure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 974-989.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558-1568.

Spelke, E. S. (1990). Principles of object perception. *Cognitive science*, *14*, 29-56.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of experimental child psychology*, *109*(1), 132-140.

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.

Swingley, D. (2010). Fast mapping and slow mapping in children's word learning. *Language learning and Development*, *6*, 179-183.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive psychology*, *54*(2), 99-132.

Téglás, E., & Bonatti, L. L. (2016). Infants anticipate probabilistic but not deterministic outcomes. *Cognition*, *157*, 227-236.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant−directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental psychology*, *39*(4), 706-716.

Turati, C., Simion, F., & Zanon, L. (2003). Newborns' perceptual categorization for closed and open geometric forms. *Infancy*, *4*(3), 309-325.

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of child language*, *5*(2), 311-326.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental science*, *7*(3), 270-276.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental science*, *10*(2), 159-164.

Wasserman, E. A., Brooks, D. I., & McMurray, B. (2015). Pigeons acquire multiple categories in parallel via associative learning: a parallel to human word learning. *Cognition*, *136*, 99-122.

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), B59-68.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends Cognitive Sciences*, *13*(6), 258-263.

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67-95.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive psychology*, *29*, 257-302.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' Ability to Learn Phonetically Similar Words: Effects of Age and Vocabulary Size. *Infancy*, *3*(1), 1-30.

Werker, J. F., Byers-Heinlein, K., & Fennell, C. T. (2009). Bilingual beginnings to learning words. *Phil. Trans. R. Soc. B*, *364*, 3649-3663.

Werker, J. F., Cohen, L. B., Lloyd, V. L., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental psychology*, *34*(6), 1289-1309.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, *7*, 49-63.

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13-and 18-month-olds. *Developmental Psychology*, *30*(4), 553-566.

Xu, F. (1999). Object individuation and object identity in infancy: the role of spatiotemporal information, object property information, and language. *Acta Psychol (Amst)*, *102*(2-3), 113-136.

Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*, *30*(2), 111-153.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223-250.

Xu, F., Cote, M., & Baker, A. (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science*, *16*(5), 372-377.

Younger, B. A., & Cohen, L. B. (1985). How infants form categories. *Psychology of Learning and Motivation*, *19*, 211-247.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37-63.

Zelazo, P. D., & Müller, U. (2002). Executive function in typical and atypical development. In I. Gouswami & U (Eds.), *Handbook of Childhood Cognitive Development* (pp. 445-469). Oxford: Blackwell.