Pragmatic word learning in monolingual and bilingually exposed children

Isabelle Lorge

Theoretical and Applied Linguistics

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

St John's College

May 2019

Unoriginally, I dedicate this thesis to my parents, without whom, after all, I would not be here (among many, many other things).

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 80,000 words including appendices, footnotes, tables and figures.

Isabelle Lorge

May 2019

Abstract

Pragmatic word learning in monolingual and bilingually exposed children

Isabelle Lorge

Humans are highly adaptable to a variety of challenging situations, as shown for example by increased echolocation abilities in the visually impaired (Schenkman & Nilsson, 2010). Multilingual input and interactions arguably create a particularly demanding environment, with added complexity and variation in the linguistic signal, a higher risk of communication failures and an increased amount of word forms to acquire. Despite this, and a lesser ability to rely on mutual exclusivity, bilingual children are able to quickly acquire a similar, and often greater vocabulary than their monolingual peers (De Houwer, Bornstein, & Putnick, 2014; Umbel, Pearson, Fernández, & Oller, 1992). A range of studies investigating attention to socio-pragmatic speaker cues found increased reliance on speaker cues in bilinguals (Colunga, Brojde, & Ahmed, 2012; Yow & Markman, 2011a, 2015). However, since these studies involved ignoring another conflicting cue, the results could have been related to inhibitory skills or to better attention to speaker generally, rather than to pragmatic inference per se, which relies on reasoning about communicative intentions.

In five separate studies, we investigated the ability of a first (n=270, range=4;1-6;2, mean age=5;3) and second (n=120, range=4;0-5;11, mean age=5;5) sample of monolingual and bilingually exposed children to use pragmatic cues to learn the

meaning of a novel word in five different tasks where success could not be achieved by ignoring a salient cue. The tasks were: contrastive inference with prosodic stress, inference based on relative frequency of a referent, ostensive teaching of a subordinate category, ostensive teaching of an action word, and use of emotional affect. We found several developmental effects, and bilinguals to be more adult-like and to significantly outperform monolinguals (compared to a baseline control condition) in all tasks which involved reasoning about communicative intentions (or why the cue was provided, i.e., the first four tasks) but not when word referent mapping could be achieved without pragmatic reasoning (directly mapping emotional valence to referent valence, i.e., fifth task).

We conclude that this thesis provides evidence for differences in the processing of pragmatic cues by bilingual and monolingual children which are not due solely to better inhibitory skills or to a general sensitivity to social cues such as prosody, eye gaze and pointing, but to performing true pragmatic inference by reasoning about communicative intentions in the context of word learning. In addition, we believe a distinction needs to be made between using social cues and reasoning about intentions, which might help provide insights about separate developmental timelines for exerting different types of pragmatic competence, with early abilities demonstrated by the bilingually exposed, particularly in acquisition contexts.

Acknowledgements

Coming to this part, one pictures it as some form of final, sophisticated torture at the end of a lengthy, eventful, yet incredibly rewarding marathon. Who to thank? In which order? Here is my best try.

First and foremost, of course, my incredibly patient, optimistic, level-headed and thoroughly wise supervisor, Dr. Napoleon Katsos, who endured four years of stubbornness, peaks of enthusiasm and wheel-rediscovering, along with periods where the end seemed nowhere in sight, with incredible good humour, positivity and diplomacy, somehow always managing to steer and channel this energy into something reasonable, and giving me much-needed perspective. I can never thank him enough for his invaluable mentoring and guidance throughout the length of both a Masters and a PhD.

Next comes the amazing group of people which constituted, successively, an experimental pragmatics 'cluster', reading group, informal reunion, and now finally officially baptised and rising 'GibKat' lab, under the direction of Napoleon Katsos and Jenny Gibson, and included, over the years, my colleagues Elspeth Wilson, Dimitra Lazaridou-Chatzigoga, Jessica Soltys, Kyriakos Antoniou, Alma Veenstra, Özge Özturk, Draško Kašćelan, Curtis Sharma, Katie Howard, Sarah Crockford, and many others. I am extremely grateful for the incredibly varied and fascinating discussions we had over the years and for their support in giving me feedback about both the intellectual and experimental aspects of my work. I would particularly like to thank Özge Özturk, Elspeth Wilson and Lisa-Maria Mueller, who agreed to involve me in the Avobe project which turned out to constitute one of my main experimental grounds, and Katie Howard for lending me her voice for audio recordings, and her whole person for video recordings. I am very grateful to the St Jerome and Norbury school staff in Harrow, Newnham Croft, St John's, St Matthews and Mayfield in Cambridge, as well as all the parents and children who agreed to take part in the studies.

I would also like to thank my former supervisor Mikhail Kissine for introducing me to the magic of linguistics and pragmatics in the first place, and the laboratory from the Université Libre de Bruxelles, with whom the GibKat lab has always had a friendly and fruitful relationship. This is in great part thanks to the Wiener-Anspach foundation, who also gave me the opportunity to come to Cambridge for my MPhil, and without whom this whole adventure would never have happened.

Finally, a thesis being the lengthy marathon we mentioned, with its ups and downs, expected and less expected bits, and generally a formative but challenging experience, I would not have dreamt of ever reaching the end of it without the emotional and affectionate support (coming in all kinds of forms, and always when needed) from my friends and family, and the financial and community support of my college, St John's.

Table of contents

List of figures

List of tables

Chapter 1

Pragmatic inferences

'When I use a word,' Humpty Dumpty said in rather a scornful tone, 'it means just what I choose it to mean — neither more nor less.' 'The question is,' said Alice,'whether you can make words mean so many different things.'

— Lewis Carroll

Glossary

Word learning cues are signals used to help determine the meaning of a novel word. Social cues are non-linguistic cues provided by a speaker for the purpose of communication.

Perceptual cues are cues linked to the perceptual characteristics (e.g., shape, colour, etc.) of an object considered as a potential candidate for a novel word.

Prosodic cues are cues linked to the speaker's voice inflexions, such as variations in pitch (ie., fundamental frequency), emotional affect. or other acoustic features (ie., volume, rhythm etc.).

Lexical or prosodic stress is the emphasised accent placed on a word or part of a word.

Emotional affect is the tone of voice typically indicative of a speaker's emotional state (ie., 'sad', 'happy' or 'fearful' tone).

Semantic cues are cues provided by the content of words, ie., their 'dictionary' meaning.

Inhibitory control refers to a subset of executive functions, namely the ability to ignore or suppress an irrelevant signal.

Linguistic cues are cues related to language: semantic (word content), syntactic (sentence structure), etc.

Contrastive inferences are pragmatic inferences based on assuming that a contrast in linguistic form indicates a contrast in meaning, for example using 'the red cup' instead of 'the cup' to describe an object potentially indicates that there are two objects which need to be distinguished.

A frequency inference is an inference based on the contextual frequency of a referent. For example, when hearing a novel word such as 'the dax'.

Ostension is emphasis placed on an action so that this action is noticed by an intended recipient. Ostensive teaching is the transmission of generic knowledge (eg., facts, language or skills) performed in a way that intends to make this intention known to the receiver.

1.1 What is pragmatics?

If you hear or read the sentence *'John is single'*, you will be able to derive from it *'John is not married'*. This is called *entailment* and will be true by virtue of the 'dictionary' or semantic meaning of the words alone. It does not require you to have access to any other type of information. By contrast, while you might come to the same conclusion upon hearing *'John spends a lot of time with young women'*, this is not strictly entailed by word meaning. Evidence for this comes, for example, from the fact that this interpretation can be cancelled, i.e., the speaker could then add: 'But he is married'.

The additional meaning drawn from the second example is called an *implicature*, a term coined by language philosopher Paul Grice (1975). Grice was the first to try and formalise that social communication cannot be achieved by means of semantic meaning and logical rules alone. To derive the second meaning you had to take into account your socio-cultural knowledge about the usual or socially acceptable behaviour of single versus married people, and potentially what you know about the speaker and their goals in sharing with you this piece of information.

Pragmatics is the study of language in context, and pragmatic inferences, or implicatures, are meanings derived by using contextual information. Context includes anything and everything that does not pertain to semantic or grammatical meaning, i.e., world knowledge, previous discourse, common ground/history between interlocutors, prosody, physical/visual environment, etc.

1.2 The Cooperative Principle

According to Grice's theory, implicatures are derived by means of the **Cooperative Principle**, that is, by assuming that the speaker is being cooperative in communicating. In this view, linguistic behaviour should generally follow certain rules or conventions governing linguistic communication, which Grice calls 'maxims': **Quality** (be truthful), **Quantity** (be informative), **Manner** (be clear) and **Relation** (be relevant). Speaker utterances can then be assumed to be cooperative even when they appear to be breaking the maxims, as in the case of metaphor and irony (*John is a pig'*; *'At least the weather is nice'* uttered in pouring rain), underinformativeness (*'This student is* *punctual and polite. Sincerely, X'* as recommendation), convoluted phrasing (*'She vocally produced notes that resembled the tune of Au Clair de la Lune'*) or unrelated contributions (*'I have work to do'* in answer to someone asking if you are planning to go to tonight's party). Under this assumption of cooperativeness, the hearer can then use context to derive the intended meaning (e.g., world knowledge about the stereotypical characteristics of pigs, presence of pouring rain, incompatibility of work and leisure, assumptions about speaker goals in writing a brief recommendation, or refraining from using the simpler expression *'she sung'*, etc.)

1.3 Relevance Theory and pragmatic trends

Consider the following sentence, uttered by a mother after her child has fallen and injured their knee (Bach, 1994):

(1) 'You're not going to die'.

Without pragmatics, this statement is doubly irrelevant, first because it seems plainly untruthful (unless the child happens to be immortal), second because it is unclear what the *intention* is in providing this particular piece of information. In the first instance, the meaning of some of the words has to be further specified using contextual information: '*You* [the child] *are not going to die* [today/from this cut]'. The hearer then again makes use of contextual information to infer the *goal* of the utterance (e.g., get the child to stop making a fuss).

One might wonder where semantics stop and pragmatics begin, or how much/how often context contributes to linguistic meaning. Grice's view, and the view of the Neo-Griceans who adapted the original theory by merging some of the principles together (e.g., L. Horn, 1984; Levinson, 1987) was to see pragmatics as an additional system complementing semantics which would enter into action when the autonomous semantic system failed or proved insufficient, and would help 'complete' linguistic meaning. By contrast, Relevance Theory (Sperber & Wilson, 1986) goes back in some way to the core of the Cooperative Principle as originally formulated by Grice: *'Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged'* (1975, p.45). In this view, cooperation, and social communication in general, boils down to being 'as relevant as required'. For the speaker, this means providing a minimal amount of linguistic signal that is enough to be reliably interpreted in the way intended, assuming the hearer is using the available contextual cues.

1.4 Aims and hypotheses

The purpose of the current thesis is to investigate the way bilingual and monolingual preschoolers learn words through pragmatic inferences. By 'bilingual', we mean anyone who is regularly and frequently exposed to two or more languages in their daily lives. In this view, the closest definition could be that of Grosjean: 'the regular use of two or more languages (or dialects), and bilinguals are those people who use two or more languages (or dialects) in their everyday lives' (Grosjean, 2008). Furthermore, we define as 'bilingually exposed' those who are regularly exposed to two or more languages, whether or not they have productive fluency in those languages. That is, we use a criterion of exposure rather than amount or fluency for including participants in our experiments, while explicitly stating the type of bilingualism when citing other works. In this way, we refer to early (prebubescent) or late bilingualism, and also mention emergent bilingualism, that is the level of bilingualism at the very beginning of learning a new language (i.e., when exposure has been a few months/less than a year). We discuss the inclusion criteria for studies more in depth in Section 2.2.2 and mention works on emergent bilingualism in Section 2.3.1.

The reason for undertaking such research is twofold. First, this thesis seeks to provide insight into an unresolved puzzle, namely bilingual children's ability to efficiently learn a greater amount of vocabulary than their monolingual peers within a similar time frame and, as we will later see, while potentially being at a disadvantage in the strategies they can use for that purpose. Second, we wanted to explore a potential advantage in socio-pragmatic abilities which could help solve the puzzle but whose nature has not been precisely defined in past research. Indeed, while some studies have provided evidence for this advantage (Yow et al., 2017; Yow & Markman, 2011a), several separate types of skills could account for these effects, and need to be distinguished.

One possibility is that this advantage is purely due to enhanced executive control skills, that is the cognitive processes used to monitor behaviour when intending to achieve specific tasks and goals (Green, 1998). Another possibility is that it results from a bilingual attentional bias towards social or speaker cues. A third final possibility resides in better pragmatic competence, that is the ability to reason about communicative intentions through cues. This thesis aims at exploring these different possibilities through a series of experiments.

Previous research has mainly focused on examining bilingual children's use of non-linguistic social cues such as eye gaze and pointing gestures (Yow, 2015; Yow & Markman, 2011b, 2016) or pitching perceptual or prosodic cues against semantic ones (Champoux-Larsson & Dylman, 2018; Colunga et al., 2012; Yow & Markman, 2011a, 2015) which arguably assesses attention to social cues (point, gaze, affect, etc.) or inhibitory control rather than pragmatic competence per se. The use of linguistic cues for contrastive and frequency inferences, emotional affect and prosodic stress in the absence of conflicting semantics, and non-linguistic ostensive teaching have not been investigated. In the studies mentioned above, better performance could have been achieved through better inhibitory control, ignoring the irrelevant semantic of perceptual information in favour of the point or gesture without having to reason about the intention behind the gesture. Alternatively, it could also have been the result of increased attention towards social or speaker cues. However, reasoning about communicative intention was not technically required, since the point or eye gaze had the effect of making a referent more salient than the others, thus salience alone could be used. In contrast, in our tasks the different conditions did not directly emphasise a referent above the others. Rather, the learners had to reason about the intention behind the cue that was provided in order to pick the correct referent.

The structure of the thesis is as follows.

In Chapter 2, we review the literature on word learning and on bilingualism and provide some evidence for a solution to the bilingual lexicon puzzle by postulating enhanced pragmatic abilities.

In Chapter 3, we examine the use of prosodic cues for fast mapping (one instance word referent disambiguation), contrasting the use of one versus two cues relying on overinformativeness and results in a task which relies on intention reasoning (contrastive inference) against the results in another task where it is not required based on emotional affect.

In Chapter 4 we examine the use of one word learning cue in the context underinformativeness in another intention reasoning task (frequency inference).

In Chapter 5 we examine two tasks relying on ostensive teaching (a subordinate category and an action word), intention reasoning and extending the newly acquired word to other instances.

In Chapter 6 we draw conclusions from our results and examine the impact of our research on the fields of pragmatic inference and bilingualism. We also conduct a quantitative analysis of our results in the context of a Bayesian framework of pragmatic inference.

If the previously found bilingual advantage is purely the result of heightened attention to social cues or greater inhibitory control, there should be no differences in performance between bilingual and monolingual subjects in our tasks. If, on the other hand, it is the result of a better ability to make pragmatic inferences, we would expect bilinguals to outperform monolinguals, even when inhibitory control alone could not have lead to better performance, and even when less cues were available (as in our first experiment), or extension had to be performed (as in our fifth and sixth experiments) but not when pragmatic inference was not required (as in our second experiment). We hypothesise that this difference might be most accurately captured in word learning paradigms since, as we previously discussed, vocabulary acquisition constitutes one of the major bilingual challenges and arguably the domain where this advantage would be most called for. We would further expect children to perform better with two available cues than one, better in ostensive than non-ostensive contexts, and better when relying on underinformativeness than overinformativeness (as the former case is generally penalised more strongly). Finally, we would expect children to perform less well than adults, indicating a developmental effect, as our tasks were chosen to be challenging so as to avoid ceiling effects.

Chapter 2

The bilingual word learning puzzle

You know, there are many people in the country today who, through no fault of their own, are sane. Some of them were born sane. Some of them became sane later in their lives.

— Monty Python

2.1 A gavagaiesque problem

In the following sections, we investigate the process of word learning, first by presenting three theoretical views on the phenomenon (the *constraints*, *social* and *associative* accounts of word learning) and discussing evidence in favour and against each of the views. We then discuss the need to separate the different steps constitutive of the process of word learning (*fast-mapping* or *referent disambiguation*, *extension* and *retention*). Finally, we present a computational approach to pragmatic inference (the Rational Speech Act Framework, or RSA) which proposes a Bayesian model for pragmatic word learning and informativeness reasoning.

2.1.1 Theories of word learning

A truly amazing feat which can nevertheless be witnessed on a daily basis in infants and young children is their ability to acquire and become proficient in a language. The learning of the very first words of a language often appears as a task of almost insurmountable difficulty, as Quine (1960) famous 'gavagai' example demonstrates: if a speaker of an unknown language points towards a rabbit scurrying by and utters '*gavagai*' how do we know what part of reality is being labelled: the rabbit, a leg, furriness, the event of rabbit running, etc.? And that is of course assuming that the hearer has already somehow managed to extract the word from the speech stream. This 'indeterminacy problem' pointed out by Quine is the main obstacle that has to be overcome for the learning to take place. Children nevertheless benefit from both a variety of (external) sources of information and several (internal) tools and skills to achieve this.

Three main sources of information help narrow down the hypothesis space of possible referents: cross-situational information (gathered through computing referent and label co-occurrences across different learning instances), social ('speaker-related') cues and linguistic (semantic and syntactic) cues. Children can efficiently use one type of cue when not all are available: they can learn without any linguistic support (as when learning their very first words) or on the contrary using only linguistic information (i.e., in the case of more abstract or functional words, Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005), they can learn without physical co-presence of the referents (as when blind or learning labels for absent referents, Bloom, 2000), using only speaker-related and linguistic information, but also without social interaction or even a speaker being present (Baron-Cohen, 1997; Scofield & Behrend, 2011; Smith & Yu, 2008).

Clearly, children will succeed in learning words in very diverse contexts and using a variety of cues, either in combination or in isolation. Less straightforward are the processes leading to referent identification and the type of abilities recruited or needed for it to take place, which, contrary to cues, cannot be observed directly. We know that the child has at her disposal several cognitive tools that can be drawn upon for word learning. Those can be grouped into three main types: associative skills, conceptual biases and inductive abilities. How do we make use of these skills in a typical word learning situation? Take Hirsh-Pasek, Michnick Golinkoff, and Hollich (1999)'s example, where a very busy host asks her guest to *'get the caponata'* (gesturing vaguely towards the refrigerator). There are several ways to reconstruct the process of referent disambiguation. If this is the first time the guest hears the word (i.e., no previous cross-situational information is available) the first cue to be used will be linguistic knowledge (if there is any). In the *caponata* example, the hearer is an adult and knows the syntax of the language and meaning of the other words. This tells us the *caponata* is 'something that you can get'. Some researchers have made a case for a crucial role of 'syntactic bootstrapping' in word learning (e.g., Gleitman et al., 2005). It is likely that different parts of speech (verbs, prepositions, etc.) will be acquired in different ways . Moreover, concreteness will play a role, with more concrete terms benefiting more from the observational context and more abstract terms from the linguistic context/the sentence structure (Gleitman et al., 2005). However, in most cases and particularly in the early days, syntax is likely to be a relatively minor cue, and purely syntactic information has been shown to be mostly uninformative for word learning (Ambridge & Lieven, 2011). Once the learner has used the linguistic information, she is left with everything else: the physical context, speaker-related information (the common ground, the gesture) and, most importantly, the internal tools or cognitive abilities that can help her exploit the information available to find the correct referent of the novel label. In all likelihood, the guest will follow the gesture towards the fridge, which happens to fit a common ground goal of setting the dinner table, open the latter and choose the unknown Italian dish, thereby learning the new word by linking the label to its referent. What are the processes that lead to the successful identification of the referent for the novel label? Learners might benefit from a set of higher-level conceptual biases or constraints to help them with the mapping task: children tend to assume that a novel word labels a whole object rather than a part, property or action (Markman, 1990; Mervis, 1987) and that it applies to something they do not yet have a name for (Markman, 1987; Markman & Wachtel, 1988; Merriman, Bowman, & MacWhinney, 1989) . In the case described above, assuming that there was only one unfamiliar dish in the fridge, the guest could have made use of these constraints to find what caponata referred to. Alternatively, children from a very young age (about 10-12 months of age) can use social cues such as gaze, gestures, physical stance, and speaker orientation to identify the speaker's focus of attention (Tomasello, Moore, & Dunham, 1995) and, possibly, make inferences about the intended referent of a novel word that are informed by the shared common ground and hypotheses about the speaker's intentions (Theory of Mind) (Clark, 2009). Thus the caponata case could be resolved by applying inductive skills to speaker-related information: the gesture can be interpreted as expressing a desire for something in the fridge, the common ground suggests it is a dish for dinner, and the use of an unfamiliar word suggests an unfamiliar referent (otherwise the speaker would most likely have used a familiar word). A final possibility is that the learner does not require any higher-level abilities (i.e., a system of lexical constraints or some type of social cognition/Theory of Mind) but is relying exclusively on lower-level mechanisms such as salience and associative skills to disambiguate the referent and link it to its label. In our example, the salience of the target referent will be higher than that of other possible referents, due to the pointing gesture, its whole object status (a whole object is more salient than a part since its parts all move together against a ground, Clark, 2009) and its novelty.

The first hypothesis about the processes involved in recovering the referent of a novel word is the **constraints account**, which postulates that children possess a set of (lexical) constraints or conceptual biases that allows them to narrow down the space of possible referents. Confronted with the need to disambiguate the referent of a novel word, empirical evidence shows that the child will usually tend to assume from very early on that the new word labels an object rather than an action or a property, that it is a whole object (whole-object assumption) and that it applies to something she does not yet have a name for (mutual exclusivity) (Markman & Wachtel, 1988). More specifically, mutual exclusivity is the tendency to map a novel word to a novel referent that has no known label rather than to a familiar one that already has a label (e.g., assuming that 'dax' refers to an unknown object rather than a cup). This has been hypothesised as being a principle resulting from assuming that every object has only one name (Markman & Wachtel, 1988), that no two words can have the same meaning (Clark, 1987) or that each object has to have a label (Golinkoff, Mervis, & Hirsh-Pasek, 1994). Another, more pragmatic possibility is reasoning that the speaker shares conventional linguistic knowledge with the hearer, knows the familiar word for the familiar object, knows that the hearer knows it too, and if wanting to be communicatively efficient would have referred to the known object with the known word (Diesendruck & Markson, 2001). One last, more recently suggested option is that this 'principle' is not a principle at all, a bias present from the start or universal in all learners, but simply tends to be the result of ad hoc language experience, continuous exposure to a system which (at least in the early stages) tends to roughly stick to a rule of one-to-one mappings, and the realisation for the child that this strategy results in a correct interpretation of labels (i.e., no communicative breakdowns) in most cases.

In terms of extension, the child will assume that it can be extended to other objects (extendibility) on a taxonomic basis (taxonomic assumption) at a basic level/the same level as the other words that have been introduced in the same manner (basic-level and equal detail assumptions)(Markman & Hutchinson, 1984; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Shipley & Kuhn, 1983). Of course, these constraints will need to be overruled in some cases for learning to take place (i.e., when learning that superordinate terms such as 'fruit' can label an object they already have a name for, e.g., 'banana'), in this sense they are usually not presented as rules but as probabilistic constraints: they help the child determine what is most likely to be referred to (Golinkoff et al., 1999). However, proponents of this account have not been clear about whether these constraints are innate and/or domain-specific (i.e., restricted to word learning) and why they should be (Ambridge & Lieven, 2011) or, if they are 'default constraints', when and where they would start to take effect/disappear (Clark, 2009). Furthermore, they seem to be inconsistent with data on spontaneous speech in children (Clark, 1997; Clark & Grossman, 1998) and some effects predicted by this system (e.g., underextension errors, such as applying the word *it* to shoes only) have not been found (Tomasello, 2003). Finally, if these constraints are neither innate, domain-specific, or covering all possible wrong assumptions about naming (e.g., assuming that names for objects change from one day to the other), it is difficult to see how it is still relevant to postulate them, instead of simply explaining the biases found in the empirical evidence through general social and/or attentional mechanisms (Deák, Flom, & Pick, 2000). As we have mentioned, a whole object is more salient because all its parts move together against a ground (Clark, 2009), objects can be grouped by similarity, often according to a salient feature (e.g., shape, giving rise to the shape bias, which might then lead to taxonomic categorisation, as we will see), new words will often refer to a basic-level category (e.g., cat), because they have more internal similarities than superordinate categories (e.g., animal) and contrast more with other basic-level categories (e.g., cat vs. dog) than subordinate categories (e.g., labrador vs. spaniel) (Tversky & Hemenway, 1984), and finally children will use contrasting words to match contrasting objects or parts (mutual exclusivity). All of these constraints thus seem like they could be the product of general perceptual/attentional processes, such as salience (Samuelson & Smith, 1998).

The second theory offering an explanation for the process through which we disambiguate the referent of a novel word is the **socio-pragmatic account**. According to proponents of this account, the narrowing of the hypothesis space of possible referents for a new word comes from interpreting the speaker's communicative intentions (e.g., Clark, 2009; Tomasello, 2003). To achieve this, the speaker first needs to establish joint focus of attention with the child through diverse means: gestures, gaze, orientation, facial expressions, prosody, etc. (Clark, 2009) and the child will use these cues and what she knows about other people's communicative intentions in general to infer the meaning that the speaker has in mind for a particular word. Some researchers have indeed suggested that social interaction is necessary for word learning: according to them, lexical acquisition could not happen from mere exposure, for example through television (Kuhl, 2007). Following this view, the guest searching for the caponata among the other objects in the fridge would not look to fill in a gap in her lexicon (i.e., 'I do not have a name for this object yet so I will apply this new label to it') as the constraints account would predict, but rather would make inferences under the assumption that the speaker is observing two major principles in her use of language: the principle of conventionality ('speakers use the conventional form in use among the community to convey their meaning') and, as its corollary, the principle of contrast ('speakers using a different form wish to convey a different meaning'). Thus, hearing the unknown word, the guest from our example might try and make an inference about the speaker's communicative intentions using these two principles, i.e., 'if the speaker had had in mind any of the familiar objects in the fridge they would have used the familiar word for it, so this novel word must refer to the unfamiliar dish'. Children do appear to make use of social information in word learning situations when it is available and, as a result, avoid mapping a new label to a referent that is not in the speaker's focus of attention: e.g. the toy they are playing with and not the one the speaker is playing with, an object that is novel for them but not for the speaker, an object that is physically present when the novel word is uttered while the target referent is not (Baldwin, 1993; Tomasello & Kruger, 1992; Tomasello, Strosberg, & Akhtar, 1996). This evidence suggests that, even though children might start applying labels by relying exclusively on perceptual features and simple associating skills, they rapidly start to use speaker-related information to override salience. A study by Moore, Angelopoulos, and Bennett (1999) showed that 12 month-old children do not follow eye gaze on a 'boring' object compared to an 'interesting' (lit up and spinning) one, but that by 19 months they will associate a novel word with the 'boring' object the speaker is looking at rather than the lit up, spinning one. Of course, the case might be made that this simply means that from a certain age social cues such as eye gaze will be more salient to the child than perceptual features such as colour or movement. Indeed, social cognition and joint attention do not appear to be crucially necessary for word learning. Studies such as Scofield and Behrend's (2011) not only show that learning is successful without intersubjective awareness, joint focus, or indeed a speaker being present at all, but also that in the first two cases (and also in the case where the focus of the speaker is ambiguous), the success rate is not significantly impaired compared to a word learning instance where speaker and child have attained joined attention. It has been found that at least some word learning could occur if children were presented with novel objects simultaneously with a novel word uttered in a disembodied voice

(Smith & Yu, 2008). Infants as young as 6 months are also able to recognise the meaning of very familiar words (Tincoff, 2001), even though socio-pragmatic skills do not seem to be used until 10 or 12 months of age, which is when joint attention can be achieved (Bloom, 2000) and children with Autism Spectrum Disorders (ASD) having very reduced social skills do succeed at learning words (Baron-Cohen et al., 1997). Similarly, if mutual exclusivity is the result of a socio-pragmatic process (i.e., trying to infer the speaker's communicative intentions), it is difficult to understand why it would lead children to ignore very strong social cues such as pointing and gazing when choosing between two potential referents for a new label (Jaswal & Hansen, 2006). While differing in the perspective they take on the process of word learning, both the constraint-based and the socio-pragmatic account have in common that they postulate the need of higher-level abilities (a system of lexical constraints/some kind of inductive ability or Theory of Mind for interpreting the intentions and actions of others) for word learning. However, research in the past years has struggled to find higher-level processes or abilities without which word learning would not be possible.

The **associative learning account** suggests that no abilities of this type are in fact required, and that learning the meaning of words can be reduced to associative learning, where the indeterminacy problem is solved using only perceptual and attentional information (such as salience) and/or statistics (by tracking label and referent co-occurrences) (Colunga & Smith, 2005; Samuelson & Smith, 1998; Wasserman, Brooks, & McMurray, 2015). Many studies have shown that this type of statistical word learning could be successfully achieved (e.g., Yu & Smith, 2008). Can we make do with just associative learning? Associative processes have been shown to involve much more that just 'simple' word-referent mapping or counting co-occurrences. Algorithms based on associations display effects and properties encountered in more naturalistic word learning situations, such as acceleration in the rate of learning, abstraction of generalizable dimensions or selective attention, rapid inference of new names, order-of-presentation effects, and more generally expectations created by a series of low-level factors (error factors, amount of experience with the object, appearance, spatial location, etc.) (Wasserman et al., 2015). However, there are strong indications that word learning cannot be reduced to associations. Several studies have shown that infants do not make mapping errors in cases of discrepant labelling (hearing a label when focussed on a non-target referent) (e.g., Baldwin,1993) and the vocabulary deficits linked to autism (Baron-Cohen et al., 1997) seem hard to reconcile with a theory of word learning as a mere associative process. As Waxman and Gelman (2009) put it, '*words do not merely associate, they refer.*' (p. 258). There is indeed a principle (not exactly a constraint) mentioned, among others, by Golinkoff, Mervis, and Hirsh-Pasek (1994), and subsumed under Clark (1988)'s principle of conventionality, that seems both to be at the foundation of word learning and not to be liable of a reduction to basic perceptual or attentional mechanisms: the principle of reference, i.e., as Golinkoff and colleagues put it, the knowledge that 'things have a name'. This does not need be an innate principle: it is possible, even likely, that children in the very first months of their life learn associations between sounds and objects or events the way dogs do, which is in a *goes with* manner (the bell goes with the door opening, the ring goes with the telephone and the word *daddy* goes with that particular person) (Golinkoff et al., 1999) and only later gain the insight that things have a name and that a word *stands for* an object (or rather, in the case of a common noun and if the child understands extendibility, a type of object). However, how the human child comes to experience that insight and shift to a 'smart' associative functioning is quite mysterious.

2.1.2 Fast mapping vs. learning

These different accounts of word learning (i.e., accounts that involve lower-level mechanisms and accounts that involve higher-level mechanisms) could possibly be reconciled by arguing that different phases of word learning will require different types of input and strategies. An interesting result of Horst and Samuelson (2008)was that retention of novel mappings was only made possible when followed by ostensive naming, i.e., the experimenter designating the object and saying ‘This is X’, which they attributed to counterbalancing the attention to the familiar competitors. Consistent with this, a more recent study traces back children with ASD’s difficulties in learning vocabulary to their failure to use feedback provided for retention (while they perform as well as normally developing children in the disambiguation task/fast mapping) (Adams-Bedford, Wallis, & Backus, 2013). This stresses the importance of separating the mechanism of first linking the word to the referenced object (whether through simple association or through higher-order inference) and the actual creation of a permanent new entry in the lexicon. These are two different steps, with one leading much less than systematically to the other. Horst and Samuelson (2008) showed that two-year-olds were very efficient at selecting a referent in a disambiguation task, but were not retaining any of the mappings that had been performed (unless ostensive naming was added). In Bion, Borovsky, and Fernald (2013), the number of (familiar) competitors in an instance of referent selection did not influence fast mapping, but negatively impacted retention for these mappings, which they attributed to the child’s attention being distracted from the novel object by the familiar competitors. It is thus possible that two different processes (probabilistic constraint-satisfaction mechanism and associative/connectionist learning) account for these different phases of word learning, requiring different resources, operating at different timescales and being interrelated without necessarily being causally linked. If the number of (familiar) competitors present during referent selection can be argued to impair retention (providing that the child indeed attends to the familiar competitors each in turn and does not 'scan' the room for a novel object, as Horst, Scott, & Pollard, 2010 hypothesise), a rich environment with diverse elements has been proven many times to enhance memorisation by acting as a retrieval cue (Dautriche & Chemla, 2014). Indeed, in Zosh, Brinster, and Halberda (2013)'s study, children's retention of mappings created through inference in a disambiguation task (i.e., with one familiar competitor: '*point at the lorp*') were significantly better than mappings resulting from direct instruction (only one object present: '*this is a lorp*'). While many explanations could account for this result (i.e., more engagement or more motivation when the child is more active in the task, more 'processing depth', etc.), there is a possibility that the presence of a small number of competitors (one, in this case) helps remembering the mapping by acting as a retrieval cue without constituting too heavy a burden in terms of memory and attentional resources (and thus risking having the opposite effect). This would reconcile tenants of statistical associative word learning and higher-level inferential strategies: fast-mapping would constitute an instance of problem solving through a socio-pragmatic or exclusion inference (ideally with an adequate number of competitors balancing indeterminacy and availability of retrieval cues), while learning and retention would happen through computing statistical regularities, with the help of a social factor.

2.1.3 A Bayesian model: the RSA framework

In the past ten years or so, several frameworks were developed which attempt to formally model pragmatic inferences using Bayesian statistics and principles derived from information and game theory. The main idea behind this approach is that speakers and listeners communicate by recursively thinking about each other's intentions. Just like in the classic prisoner's dilemma (Rapoport, Chammah, & Orwant, 1965), the choice of behaviour in a social context is influenced by the other intelligent entities involved. In this famous dilemma, two prisoners with no means of communicating are given two choices: betray the other and walk free while the other prisoner serves three years, or stay silent. If both prisoners stay silent, both serve one year ; if both speak, both serve two years. A rational individual who does not reason about the other prisoner as an intelligent being should choose to betray, since it provides a better outcome whatever the other prisoner chooses. However, if both prisoners reason about the other's best individual choice and realise this leads to both betraying (and thus both serving two years) and that the other prisoner will have come to this conclusion too, they might decide to stay silent on the assumption that the other prisoner will want to cooperate to avoid a rational equilibrium with a worse mutual outcome. This is in effect what often happens when humans are presented with this dilemma.

The prisoner's reasoning involves several steps of recursive reasoning (thinking about what the other prisoner is thinking about what the first prisoner is thinking etc.). Communication can be seen as a game where both interlocutors are seeking to maximise their cost/benefit efficiency. A speaker choosing an entirely self-centered behaviour (minimising their cost and maximising the listener's) would not say or do anything and leave all communicative cost to the listener. They would be assuming very sophisticated, almost supernatural, mind reading powers on the part of the listener (i.e., telepathy). On the other end of the spectrum, a speaker who assumed the listener not to be performing any such mind reading would have to maximise their own cost by being as explicit as possible (just like when interacting with a computer for example). Efficient communication through pragmatic inference results from searching the space of possible behaviours to reach a mutual equilibrium between these two extreme cases, where speakers are providing listeners with just enough input for them to be reasonably expected to correctly infer meaning (Franke, 2011).

While in theory this could involve infinite recursion, in practice the Rational Speech Act (RSA, Frank & Goodman, 2012) framework provides a starting point with a literal listener L0 expected to calculate probabilities for referents based on semantic meaning only.

Thus, in the classic scalar implicature example (e.g., *'Some of the students failed'*), L0 computes the probability of *some* to be 0 for non compatible meanings and 1 for all compatible meanings, equally divided between them, i.e., 0.5 for 'all' and 0.5 for 'some not all'. Similarly, L0 can compute the probability of *all* as 1 for the meaning 'all'.

A pragmatic speaker is assumed to be rational and choose an utterance to maximise benefit (or utility), i.e., the chance that the literal listener will infer the correct referent. The word *all* has a utility of 1 for the meaning 'all' whereas the word *some* would only have a utility of 0.5 for the same meaning.

Finally, a pragmatic listener reasons about a pragmatic speaker (who chose their behaviour by reasoning about the literal listener) to infer the meaning of the word or utterance used. Bayesian statistics come into play by providing the posterior probability of each alternative referent as the product (or addition, in log space) of the prior probability for the referent and the likelihood given the word used. Assuming a uniform prior (i.e., equal prior probabilities for all meanings), since the utility of *some* for conveying 'all' is half as high as that of *all*, it is half as likely to have been used to mean 'all' than 'not all'.

The extent of this recursive process is subject to debate. Indeed, studies examining the issue find little evidence for humans reasoning beyond one recursive step (corresponding to second-degree Theory of Mind, i.e., reasoning about what the interlocutor thinks you are thinking) (Degen, Franke, & Jager, 2013) and, as we discuss in the next sections, even first degree Theory of Mind might not always be required (Keysar, Barr, Balin, & Brauner, 2000) which indicates that perhaps a more minimalist Bayesian model could be proposed (Muhlstein, 2016). Indeed, the 'some not all' implicature could perhaps be arrived at by simply considering the specificity (or 'narrowness') of each hypothesis: *some* is twice as specific for the meaning 'some not all' (which can only be expressed using this word) than for 'all' (which can be expressed using a more informative word).

However, the RSA framework has successfully modelled a wide range of pragmatic phenomena, from choices of modifier (Frank & Goodman, 2012) to choices of reference level (Graf, Degen, Hawkins, & Goodman, 2016), cross-situational word learning (Goodman, Tenenbaum, & Black, 2008), *ad hoc* quantity implicatures (Stiller, Goodman, & Frank, 2015), word learning based on frequency (Frank & Goodman, 2014) and inferring properties of novel categories from referring expressions (Horowitz & Frank, 2016).

In one of their first studies of this kind, Frank and Goodman (2012) presented participants with a referential language game where they had to refer to coloured shapes (e.g., a blue square, blue circle and green square) using only one word. They collected data both on prior probabilities of referring to each of the shapes (before any words had been uttered) and on how likely listeners thought each shape was to be referred to, given a single word uttered (e.g., *'Which shape would you bet on if you heard the word 'blue"?*). The authors then built a model based on the Bayesian framework mentioned, using the combination of prior probabilities and likelihoods of each shape given the word uttered, as proportional to the utility, of informativeness of the word for each shape (e.g., 'blue' is twice as informative for the blue square as for the blue circle). They found participant's responses to be closely quantitatively matched to the predictions of the model. A similar framework was used in Frank, Goodman, and Tenenbaum (2009) to model word learning by simultaneously combining cross-situational information (as raising the prior probabilities of some referents) and reasoning about intentions, and found to outperform models which only used one source of information.

In conclusion, it seems that theories of word learning can only be successful in producing accurate models of performance when taking into account the diversity of sources that children can draw upon and combining/weighting them in a probabilistic way (a constraint-based, or cue integration approach, which we return to in later chapters), but also the different developmental timelines involved in each type of cue/source of information, and the varying weights given to these cues in each of the steps leading to the creation of a novel entry in the mental lexicon.

2.2 Adding complexity

In the following sections, we present an overview of the problem of word learning in a multilingual environment. We begin by describing how an environment involving multiple languages has a potential impact on a number of cognitive, social and linguistic aspects of development. We then outline how these differences might influence the way children growing up with two or more languages build a vocabulary in each of their languages. We finish by examining one attempt at modelling multilingual word learning using a Bayesian framework similar to that of the Rational Speech Act, which show poor performance despite some tentative adaptations.

2.2.1 The multilingual environment

While monolinguals are having to perform intricated bootstrapping to overcome a seemingly unsolvable indeterminacy problem, a substantial number of children are faced with the task of acquiring not just one, but multiple languages. These early bilinguals will need to both quickly realise the presence of multiple systems and to learn to distinguish between them. Both early and late bilinguals will learn to switch between and inhibit the languages (Green, 1998). In addition, no matter the age at which a second language is learned (this includes late L2 learners), bilinguals have to face greater complexity (syntactic, semantic, lexical, prosodic and phonological) in the linguistic input they process as well as in the communicative interactions they are involved in. For early, but also late bilingual learners, this greater complexity of input might mean that some categories (semantic, phonological, and perhaps syntactic) will tend to be built or to converge in a hybrid rather than separate fashion (Storms, Ameel, & Malt, 2015) or, in the case of number of lexical items in each language, result in some early delays (Bialystok, Luk, Peets, & Yang, 2010).

One correlate of bilingualism that has been given a lot of attention is the need to exert some type of control over the two languages to allow the bilingual speaker to switch between languages when needed and inhibit the irrelevant language while using the other one (Green, 1998). An extensive body of experimental literature has shown early bilingual children outperforming monolinguals on tasks related to executive functioning skills, such as working memory, inhibition and cognitive flexibility (Barac & Bialystok, 2011; Bialystok & Craik, 2010). Yet many studies have failed to replicate these results and the very existence of this alleged advantage has recently been questioned (Paap, Johnson, & Sawi, 2015). However, the inconsistency in the results could be due to different types of bilingualism having differential influences: earlier acquisition, higher proficiency and more frequent switching between languages are for example likely to result in a higher need for control abilities (Green, 2011). Improved skills in executive control could enhance the basic associative mechanisms underlying instances of word learning by allowing the bilinguals to retain a greater number of possible referents for a novel word and switching more easily between hypothesised referents. This advantage could be mediated by verbal short-term memory (Papagno

& Vallar, 1995) or inhibitory control (Bartolotti, Marian, Schroeder, & Shook, 2011). The cognitive advantages experienced by bilinguals could also potentially have an impact on word learning beyond simple associative mechanisms, by improving their ability to combine cues to word meanings from different sources and ignore irrelevant or contradictory information (Marian & Kaushanskaya, 2007).

Another obvious difference between bilinguals and monolinguals is that the former possess not one but two different semantic systems, which appear to exert an influence on each other at a lexical as well as a conceptual level. As a result, bilinguals appear to experience facilitation in acquiring certain types of words. For example, vocabulary checklists in simultaneous bilinguals (English-L2, aged 6-92 months) show that they learn more easily words that they have already learned in one language, giving rise to overproduction of translation equivalents (TEs) compared to hypothesised independent systems, and resulting in sparser lexicons than monolinguals' (Bilson, Yoshida, Tran, Woods, & Hills, 2015). Early sequential bilinguals in a picture-naming task (Samoan-English, aged 4.7-5.2) also learn more easily words that are conceptually closer in both languages (Hemsley, Holm, & Dodd, 2013), whereas late proficient bilinguals (English-Spanish, mean age= 24.91) are better at learning novel words with a higher level of concreteness, probably due to the fact that they share more conceptual features across languages (Kaushanskaya & Rechtzigel, 2012) . The two lexicons also appear to interact at a conceptual level. Empirical evidence indicates that activation of a lexical unit activates all conceptual features associated with it, including features from the non-target language (Colomé, 2001; Costa, Miozzo, & Caramazza, 1999). For example, in a French-English bilingual, the English word 'finger' is likely to activate information related to both fingers and toes, given that the same word (doigt) is used for both in French. In proficient late adult bilinguals, this co-activation can result in 'in the moment' semantic transfer, where a word of phrase is used in a way that is consistent with its semantic content in the non-target language (Elston-Güttler & Williams, 2008; Marian & Kaushanskaya, 2007)). Alternatively, it also appears to give rise to a deeper change in lexical representations, creating overextended or 'in-between' categories that do not correspond to either monolingual equivalent, revealing a bidirectional influence for both balanced early bilingual adults (Dutch-French, Ameel, Malt, Storms, & Van Assche, 2009; Ameel, Storms, Malt, & Sloman, 2005) and children (aged 5-14, Dutch-French, Storms et al., 2015), whereas only an influence of L1 on L2 was found in late bilingual adults (aged 18-25, Arabic-English, Gathercole & Moawad, 2010).

Early bilinguals have long been thought to have better metalinguistic skills (including word awareness, grammatical awareness and phonological awareness) than their monolingual counterparts (Ben-Zeev, 1977; Bialystok, 1988; Cummins & Mulcahy, 1978; Ricciardelli, 1992). This would result from early insights about the explicit functioning and arbitrary nature of language derived from the availability of two alternative linguistic systems. As such, bilingualism is also likely to influence metalinguistic abilities more particularly related to word learning mechanisms and principles. In terms of word learning strategies, an important point where bilinguals and monolinguals have been seen to diverge is application of mutual exclusivity. As we have seen previously, when learning words, children (and adults) tend to disambiguate between referents by applying a principle of mutual exclusivity (lexical constraints account) or, possibly, by assuming that the speaker behaves according to a principle of contrast (socio-pragmatic account). The mutual exclusivity results in a tendency to attribute a new name to a new object rather than a familiar one, e.g. for children to decide to use a new label like *zav* to refer to an (unknown) garlic press rather than to, e.g., a cup for which they already have a name (Markman & Wachtel, 1988). This principle seems not to apply between languages in simultaneous and early sequential bilinguals and to be generally weaker for bilingual than for monolingual children (3-6 years old, English-Urdu, Davidson & Tell, 2005; 2-4 years old, Danish-English, Healey & Skarabela, 2008; 2-5 years old, English-L2, Bialystok, Barac, Blaye, & Poulin-Dubois, 2010). Furthermore, this tendency seems to correlate with number of languages (Byers-Heinlein & Werker, 2009) and lexicon structure in simultaneous bilinguals (17-18 months, English-Chinese, Byers-Heinlein & Werker, 2013) with more translation equivalents (TEs) leading to a less reliable application of mutual exclusivity. However, this difference might not lie in the use of disambiguation itself but rather in the degree of acceptance of lexical overlap. A study by Kalashnikova, Mattock, and Monaghan (2015) tested these two tendencies separately in 3 to 5 year old simultaneous bilingual children (English-L2) by having them perform a classic mutual exclusivity task and another task where two labels were introduced for the same unfamiliar object, and found that bilinguals were successful in performing the first task but showed a greater acceptance of lexical overlap in the second task. A study by Rowe, Jacobson, and Saylor (2015) conducted on simultaneous bilinguals (aged 3.5-5.11, English-L2) also showed that they were more likely to accept two overlapping labels for one object given some pragmatic information to that effect. In their study, children were presented with a familiar object including a salient part (for example, a boat with a cabin), and a novel label was offered, either without any extra information in the baseline condition, (e.g., *'this is a skiff'*), with a small amount of pragmatic information encouraging the drop of mutual exclusivity constraint in a second condition (e.g., *'this is a skiff, it's a boat'*), or with a high amount of pragmatic information in a third 'rich' condition (e.g., *'this is a skiff, it's a kind of boat'*). Bilinguals were above chance in the second condition in using the label for the whole object rather than the part, whereas monolinguals needed a higher amount of information to accept two names for the same object.

Bilingual experience also appears to have an influence on the application of other principles of word learning, such as the principle of conventionality, and on patterns of categorisation. Henderson and Scott (2015) show that simultaneous bilingual infants (M=13 months, English-L2) are more conservative in their assumptions of conventionality, extending conventional meanings less both within and between languages and speakers. In their study, bilingual and monolingual children were familiarized with two speakers who both spoke the same language, or a different language, and were then presented with an habituation word learning paradigm. Statistical analyses of looking time measurements showed that bilinguals did not expect the same label to be used by speakers of the same language, and were actually surprised (looked longer) when two speakers of different languages used the same label to name a novel object. As regards categorisation, Storms et al. (2015) tested simultaneous and early sequential balanced bilinguals (aged 5-14 years, Dutch-French) in a container-naming paradigm where they had to attribute a label (e.g., bottle, flask, etc.) to different types of containers for which the label/type mappings do not systematically overlap between languages. Not only did they find, as we mentioned above, that bilinguals' categorisation patterns in both languages did not match either monolingual equivalent, but bilinguals also appeared to rely for a longer time on similarity-based categorisation, starting to produce language-specific exceptions to that principle around 14 years of age, while they appeared around age 10 in monolinguals.

Some researchers (Byers-Heinlein & Werker, 2009; Houston-Price, Caloghiris, & Raviglione, 2010) suggested that the presence of translation equivalents in the bilingual's lexicon could give rise to an early understanding that one referent can have two labels, and thus to a less systematic application of the mutual exclusivity principle. Disambiguation would then be directly influenced by lexical knowledge and more particularly by the number of *many-to-one* (inter and intra-linguistic synonyms) versus *one-to-one* mappings contained in the lexicon. This could explain both why disambiguation does not arise concurrently with word knowledge but significantly later (Halberda, 2003; Markman, Wasow, & Hansen, 2003), and why there seem to be differences in its application by multilingual infants. This could mean that there are at least two lexical profiles for bilinguals: balanced bilinguals, having more translation equivalents and thus also allowing (through a weaker use of mutual exclusivity) for more intra-linguistic synonyms (amounting to more *many-to-one* mappings and a less 'conceptually diversified' vocabulary) and dominant bilinguals, showing the opposite trend (and a maybe more 'conceptually diversified' lexicon). This increased reflective perspective on language could also lead to a different perception of other word learning principles, resulting in bilinguals making different linguistic 'bets' than monolinguals, i.e., generalising certain principles less (such as conventionality or mutual exclusivity) and other principles more (such as similarity-based categorisation).

Finally, bilinguals appear to show different sensitivities to cues for word learning and segmentation. Gervain and Werker (unpublished results)'s study using a high-amplitude sucking paradigm (HAS) shows that 7-month simultaneous bilinguals (English-L2 with a different word order) are able to use prosody to segment noun phrases, whereas monolinguals do not. Moreover, at 8 months they continue using cues such as face movements in a task of language discrimination, whereas monolinguals can use this cue to discriminate between languages at 4 and 6 months but no longer pay attention to it at later stages (Werker, 2012).

2.2.2 Building a lexicon

Can we see in bilingual lexicons evidence of differences in word learning efficiency or abilities? No matter the type of bilingualism, or indeed, multilingualism involved, a common and defining feature is a reduced amount of input per language. Children growing up in multilingual environments are by definition exposed to only a subset of the input monolingual children receive for the same language. Even in cases where the multilingual exposure is very limited (say, 5%), this is still denting on the equivalent input that would be entirely dedicated to a single language. It is, however, unlikely that a perfectly linear correlation exists between amount of exposure and vocabulary proficiency (an hour of exposure doesn't equate an hour of 'pure' input, and the quality of that input may vary widely, not to mention the subject's learning abilities). So, to what extent does this reduced input impact the bilingual's word knowledge?

Unsurprisingly, receptive and productive vocabularies of bilinguals do seem to lag behind those of their monolingual counterparts, at least at early ages. In terms of scale, the most prominent study examining this effect is that of Bialystok, Luk, et al. (2010), who collected standard vocabulary test scores data from different studies for 1,738 children aged 3 to 10 years old. Of these, 966 were classified as 'bilinguals', defined as children being schooled in English, speaking another language at home on a daily basis, and reported by parents to be fluent in both languages. This description highlights the importance of spelling out clearly the criteria for participant selection in studies that include bilingual subjects. Failing to do so leads to the risk of generalising to the whole bilingual population effects or advantages (such as cognitive advantages) that may be the prerogative of a certain type of bilingualism only.

Bilingualism can be described along three (highly correlated) main dimensions: frequency, amount and fluency. All studies agree that a child experiencing multilingual exposure that is not regular, substantial, or resulting in any degree of linguistic competence does not classify as bilingual. The most widely used entry criterion is frequency: the exposure has to be regular. The rule of thumb is generally that one or both parents speak another language than the school language (which approximately ensures daily exposure to both languages). Additionally, many studies require that a certain threshold amount of exposure be reached, generally 25% (Byers-Heinlein, 2017; Byers-Heinlein & Werker, 2009; Yow et al., 2017) or 20 hours a week (Junker

& Stockman, 2002) and that the child be described by her parents as 'fluent in both languages' (Akhtar, Menjivar, Hoicka, & Sabbagh, 2012; Bialystok, Luk, et al., 2010) or 'being able to speak and understand both languages' (Fan, Liberman, Keysar, & Kinzler, 2015). Older studies relied mainly on the frequency criterion, classifying as 'bilingual' any child who had regular exposure, even non-daily or minimal (less that 25%) to another language (Pearson, Fernandez, Lewedeg, & Oller, 1997; Pearson, Fernández, & Oller, 1993) . In more recent studies, however, there is a tendency to distinguish bilingual from 'bilingually exposed' children using criteria of both amount of exposure and fluency, and to label with the latter term a child that benefits from 'minimal' regular exposure to another language (e.g., weekly, or less than 20%) and cannot be described as fluent in that language (Akhtar et al., 2012; Fan et al., 2015; Menjivar & Akhtar, 2017). There does not appear to have been attempts to include and/or label as bilingual individuals with opposite profiles, probably because multilingualism could then not be described as enough of a part of the child's daily life to be of consequence (e.g., playing with Spanish cousins during the summer holidays) or because this profile is unlikely to be encountered in children and/or constitutes an unstable situation liable to attrition phenomena (e.g., having a high competence in French but hardly ever putting it to use).

As we said, studies such as Bialystok et al.'s (2010) which examined the lexical knowledge of bilingual children (as defined by the criteria we mentioned previously; i.e., regular exposure and fluency in more than one language) did find significantly lower levels of performance in bilinguals. However, several points regarding these results should be raised.

First, most of these studies, including Bialystok et al.'s (2010) have assessed vocabulary knowledge in only one of the child's languages, generally English, which represents only part of the 'bilingual picture' and total vocabulary knowledge of the child. Second, this disadvantage is far from being as extended as would be expected both from the reduced amount of input and the consequences of multilingual exposure in terms of available learning strategies, as we will explain later. While Bialystok et al.'s (2010) found performance to be only around 10% lower for bilinguals in English (with a mean of 96.3 for bilinguals and 106.8 for monolinguals on the PPVT), most studies which only assess English tend to find levels of receptive vocabulary for bilinguals to be about 60 to 90% of monolinguals', mostly in the upper end of this range (Byers-Heinlein & Werker, 2009; Fan et al., 2015; Yow, 2015; Yow & Markman, 2015).

What about the other language? Do children reach these normal levels of English vocabulary despite having about half the exposure of the monolinguals' (most studies average 50% exposure for each language) at dire cost in their other language? It appears not. What we learn from the few studies which examined vocabulary knowledge in both languages known by the child is that having about 60 to 90% as many words in one language as monolinguals does not mean having only about 10 to 40% in the other language (which would amount to a similar number of lexical items as monolinguals). In the vast majority of cases, bilinguals simply have more words. Umbel et al. (1992) studied 105 Miami first graders and found their performance on the PPVT to be between 69.7 (Spanish only at home) and 88.0 (English and Spanish at home), while their TVIP (Spanish equivalent of the PPVT) was on par with monolinguals' at 97. Junker and Stockman (2002) examined the productive vocabularies of ten German-English (one person/one language) bilinguals with two carefully matched monolingual groups and found that they could produce about 50% more items (approximately 300 words in total), while their conceptual vocabulary (calculated as the number of concepts for which the child knows a word in at least one language, i.e., translation equivalents referring to the same concept such as 'cat' and 'Katze' being counted as one lexical item) was the same (about 200 concepts). Scheffner Hammer, Lawrence, and

Miccio (2008) measured the receptive vocabularies of 83 Miami children in year 1 and 2 found their scores in both languages to be between 62 (for Spanish in homes where both Spanish and English were spoken) and 87 (for English in the same situation). Similarly, De Houwer et al. (2014) investigated the productive and receptive lexical knowledge of 31 French-Dutch (one person/one language) bilinguals and found that, when both languages were combined, they understood significantly (about 71%) more words than monolinguals, while their production levels, either for Dutch only or both languages combined, did not differ significantly from monolinguals'.

One exception to bilinguals knowing significantly more words than monolinguals (and the only one known by us) is Pearson et al.'s (1993) seminal study, which found levels of vocabulary in bilinguals to be approximately correlated with levels of input, i.e., about 50% of monolinguals' in both English and Spanish. However, the sample size used was small (25), some bilinguals had very unbalanced levels of exposure (less than 25%), and the results are altogether surprising: the bilinguals' conceptual vocabulary was approaching monolingual levels, which means most of the words known in either language had no translation equivalent in the other language, a situation that would be both unlikely and communicatively inefficient. Indeed, while some words are bound to be language and situation-specific (for example, restricted to the home or school environment), many words will be relevant to all communicative settings.

Even if we make the conservative assumption that Pearson et al. (1993)'s results were truly representative of their sample's skills, these are still levels of performance that are within the normal range and do not fall below what would be expected from the input, assuming no other disadvantages compared to learning in a monolingual environment. Can we actually make this assumption? Without even considering difficulties related to greater variation, multilingual input potentially results in language learners being less able to rely on some of the most widely used word learning strategies, such as mutual exclusivity, or some assumptions about conventionality.

The idea that mutual exclusivity is a direct consequence of specific language experience derives from observing that it is not available from the onset of language but later, around 17 months of age (Halberda, 2003) and that it appears to be less reliably used by bilinguals (Bialystok, Barac, Blaye, & Poulin-Dubois, 2010; Byers-Heinlein & Werker, 2009, 2013). Even more, it seems to be directly linked to the type of linguistic input the child has experienced. Byers-Heinlein and Werker (2009) studied three groups of 16 children aged 17 or 18 months old, either monolingual (no systematic exposure to another language than English), bilingual (at least 25% exposure to other language since birth, average 48% English) or trilingual (at least 20% exposure to two other languages since birth, average 47% English). They found that monolinguals significantly increased their looks towards the novel object when hearing a novel word, whereas that tendency was only marginal in bilinguals, and absent in trilinguals. This suggests that the type of language experience, but also, as demonstrated in a later study (Byers-Heinlein & Werker, 2013) the type of lexicon already acquired, and whether it contains many translation equivalents (resulting in one-to-many mappings), are directly correlated to the use of mutual exclusivity. An interesting result of their study is that, despite having similar or even lower levels of exposure and making significantly less use of mutual exclusivity, trilinguals actually outperformed bilinguals on vocabulary scores and knew about 77% of monolinguals' words, showing that they were able to learn as much if not more words without relying on this strategy. Studies have tended to suggest that multilingual exposure affects expectations about conventionality more generally: contrary to monolinguals, bilingual toddlers do not apply mutual exclusivity between languages (Byers-Heinlein et al., 2013), expect the meaning of a word not to be shared between speakers of different languages (Henderson & Scott, 2015) and bilingual infants appear not to share monolinguals' expectations of one-to-one relations between words and referents, i.e., that two labels should not refer to the same category, or the same label to two different categories (Byers-Heinlein, 2017). This begs the question of how these expectations relate to actual word learning performances; i.e., how necessary they are and what do children use if they are not using them?

2.2.3 Modelling bilingual word learning

As mentioned in section 2.1.3, some computational models have successfully used Bayesian probabilities to build computational models of pragmatic inferences, including those involved in word learning. Zinszer, Rolotti, Li, and Li (2018) recently attempted to use a similar model to predict learning with multilingual input and obtained (as would be expected) very poor performance with a strict rule of mutual exclusivity arrived at through a requirement for scarcity (i.e., encouraging smaller lexicons). However, even when spelling out mutual exclusivity in a more straightforward way (by discouraging one-to-many mappings) and allowing it to be a flexible parameter tuned by the algorithm to the type of linguistic input (monolingual or bilingual, in accordance with Byers-Heinlein et al. 2009's findings), the optimised version of the algorithm still falls very short of the performance witnessed in most children's data, learning about the same number of mappings as with monolingual input (16.75 versus 15.5), when the number of items to be learned in the bilingual version of the input is almost double (50 versus 34). As the authors agree, '*in an experimental paradigm using a comparable training period [. . .] children who receive bilingual input would learn many fewer words than their monolingual counterparts*'. As we have seen, this is not what we observe in children's data, where bilinguals generally achieve about 60 to 90% of the monolingual performance in each language, leading them to know about 50 to 70% more lexical items.

Evidence from many studies investigating vocabulary scores in bilinguals indicate that their performances are on par with monolinguals' in each language, resulting in them knowing a significantly greater amount of lexical items, despite much reduced input and being less able to rely on widely used word learning principles such as mutual exclusivity, or monolinguals' assumptions about conventionality. However, when trying to model pragmatic word learning with bilingual input, even using a flexible parameter for mutual exclusivity, performances are very poor and do not appear to replicate bilinguals' success in learning a great number of words. In addition, some studies carried on adult late (Papagno & Vallar, 1995) and early (Kaushanskaya & Marian, 2007) bilinguals have suggested a general advantage in word learning tasks, as well as better performance in learning transitional probabilities in a second language. Simultaneous bilinguals also appear to be more efficient at learning speech structures: Kovács and Mehler (2009b) tested 12 months-old infants in learning two different patterns of regularities where the structure of a trisyllabic item predicted the location of a toy and found that bilinguals were able to reliably learn two systems whereas monolinguals could only learn one. However, a recent study by Byers-Heinlein et al. (2013) using an eye-tracking paradigm to investigate associative learning in 12 and 14 months-old simultaneous bilinguals (English-L2, min 25% exposure) and monolinguals did not find any significant differences of performance between the two groups.

The question is then the following: how do bilinguals manage these impressive performances? We already mentioned some possible factors in section 2.2.1, such as cognitive advantages, conceptual transfer or better ability to use cues such as prosody to learn word transitions and segmentation patterns.

Another possibility, given that most studies show that bilinguals have a significantly greater total vocabulary than monolinguals, and that vocabulary size has been shown to be positively correlated with word learning performances, is that they are able to pull on the entirety of their lexical knowledge to help subsequent word learning and thus have an advantage over monolinguals, which might also explain the better word learning performances of bilingual adults (Kaushanskaya & Marian, 2009). However, the relationship between vocabulary size and foreign label learning is uncertain, with some studies showing a positive effect (Koenig & Woodward, 2012), while others do not (Akhtar et al., 2012; Menjivar & Akhtar, 2017). Akhtar et al. (2012) found that bilingually exposed children were able to learn foreign labels while monolinguals and bilinguals were not, but it should be noted that the bilinguals in this study were of significantly lower SES than the two other groups, and that in a subsequent study were SES levels did not significantly differ between the three groups, foreign label learning was positively correlated with level of bilingualism, with bilinguals significantly outperforming monolinguals and bilingually exposed children's performance falling in between (Menjivar & Akhtar, 2017). Performance was, however, not correlated with vocabulary scores.

In conclusion, multilingual input is a major factor in development, which has consequences on many aspects of cognitive, social and linguistic skills. As a result, constructing a lexicon might proceed differently in multilingual environments, as there are differences in the skills which can be used for this purpose and the tools or principles which can be relied on (such as the mutual exclusivity principle). Despite this, an extensive review of the literature shows that bilingual children do learn words quickly and efficiently, and do not display major or problematic gaps in their knowledge. However, current multilingual adaptations of word learning models with integration of cross-situational and social information, such as the Rational Speech Act framework, do not appear able to model these performances. Thus, some other factor needs to be taken into account to explain how children growing up bilingually adapt in order to perform efficient word learning despite a significantly different environment. We then next turn to pragmatic skills as a potential solution to this puzzle.

2.3 A pragmatic solution?

In the next sections, we outline a possible solution to the puzzle of how bilingual children manage to acquire vocabulary in an efficient way, namely by displaying an advantage in pragmatic reasoning which would arise both from additional experience with a diversity of social situations and challenging communication, and from the need to compensate a lack in vocabulary knowledge and less ability to rely on some word learning principles and strategies. We first describe research on pragmatic competence and bilingualisms, which does indicate the presence of a potential advantage. We then discuss the different types of pragmatic inferences, so as to introduce the issue of defining the precise nature of this advantage, if it exists. We finally present the aims of the current work and the three hypotheses put forward a potential socio-pragmatic advantage, focusing on the last one as our main hypothesis.

2.3.1 Pragmatic skills and bilingualism

One possibility explaining the puzzle of bilinguals' impressive success in building a large vocabulary is that they use alternative strategies for word learning and rely on more sophisticated socio-pragmatic skills fostered by their learning environment to keep pace with their monolingual counterparts, rather than on word learning heuristics such as mutual exclusivity (Byers-Heinlein, 2017).

Children growing up in bilingual environments have to face additional challenges in communicative interactions, which arise from having to identify the linguistic profile of any given interlocutor (monolingual in language A, monolingual in language

B, bilingual without code-switching, bilingual who engages in code-switching, etc.) and result in a higher risk of communication failures. Evidence shows that children confronted to a more challenging, i.e. more varied and complex input (more phonetically diversified, or provided by different adults) will map sounds to meanings more efficiently (Graf Estes & Hurley, 2013) and attend more to the relevant aspects of a word-learning task (Shneidman, Buresh, Shimpi, Knight-Schwarz, & Woodward, 2009). This demonstrates that social experience can influence how children attend to and learn from others (Shneidman, Gweon, Schulz, & Woodward, 2016) and that they will learn to allocate attention more efficiently when faced with more uncertainty in their environment. In line with this, bilingual toddlers (but not bidialectals, who might experience less comprehension or communication issues) appear to be better at detecting and attempting to repair communication misunderstandings (Wermelinger, Gampe, & Daum, 2017) and exposure to code-switching (creating more linguistic complexity and uncertainty) seems to increase attention to social cues and speakers in both monolingual and bilingual children (Yow & Markman, 2016). Similarly, bilinguals have consistently displayed enhanced Theory of Mind (Goetz, 2003) and perspective-taking abilities (Greenberg, Bellana, & Bialystok, 2013; Rubio-Fernández & Glucksberg, 2012). Unlike the potential early delay in vocabulary knowledge, this higher risk of communication failure applies to both early and late bilinguals, and even to L2 learners. Yet, research on what is sometimes referred to as 'emergent bilingualism' (limited bilingual exposure) has mostly focused on finding a threshold for apparent cognitive benefits, in 9 month-old infants (Kovács & Mehler, 2009a), children attending immersion programs (Bialystok & Barac, 2012) and young adults learning a second language (M. D. Sullivan, Janus, Moreno, Astheimer, & Bialystok, 2014). There have also been attempts to characterise the cognitive differences between the effects of bilingualism and bidialectalism (Ross & Melinger, 2017).

The few studies who did use bilingually exposed children (i.e., who might not have fluent productive ability) have found similar results as with productive bilinguals for perspective-taking (Fan et al., 2015, Liberman et al., 2017) and for word learning (Menjivar & Akhtar, 2017) as well as both bilingual adults and L2 learners adapting their speech to a greater extent to avoid communication failures when talking to a child or non-native speaker (Lorge & Katsos, 2018) . As we have seen, bilinguals have to learn (and successfully do learn) a greater amount of total words while being less able to rely on word learning strategies such as mutual exclusivity and also have to communicate with less vocabulary at their disposal in each of their languages. In addition, the added complexity and uncertainty in their learning environment appears to make them more socially and communicatively aware. Could this situation result in enhanced pragmatic abilities and could these abilities (as has been suggested, Byers-Heinlein, 2017) help them compensate the obstacles they face to achieve efficient vocabulary learning?

A few studies have investigated potentially enhanced pragmatic abilities in bilinguals. However, most of these tasks had a strong metalinguistic component, which makes it harder to know precisely which skills are being assessed, or were conducted with older, school-age children well after the age where efficient word-learning skills would be assumed to be most needed. In three studies (Siegal, Iozzi, & Surian, 2009; Siegal, Matsuo, Pond, & Otsu, 2007; Siegal et al., 2010), Siegal and colleagues conducted acceptability judgments on scalar implicatures and presented children aged 3-6 years old with statements from a CVT (Conversational Violations Test) uttered by two dolls (e.g., 'What did you get for your birthday? 'A present') asking them which of the two dolls had said something 'silly or rude'. Antoniou, Veenstra, Kissine, and Katsos (2018) and Antoniou and Katsos (2017) used a combination of acceptability judgments (which can sometimes require extensive world knowledge) and picture choices, however even in the latter cases some metalinguistic component might still have biased the basis for answering. For example, in the paradigm adapted from Kronmüller, Morisseau, and Noveck (2014), a character has a hidden and visible card and refers to the visible item as *'the open window'*. The participant is then asked 'what do you think was on Martijn's second card?' which arguably both requires an extra step of reasoning but also drives the participant to reason about the use of this question as well as the use of the modifier in 'open window').

Other studies have investigated the use of 'socio-pragmatic cues' for word learning and referential resolution in bilingual children. For example, a study by (Yow & Markman, 2015) suggests that simultaneous bilinguals (aged 3.46-3.98, English-L2, min 30% exposure in non-dominant language) are better than monolinguals at combining speaker's cues to identify a novel words' referent. In a replication of Nurmsoo and Bloom (2008)'s paradigm, they presented 3-year old bilingual and monolingual children with a word-learning task where they saw two objects while the experimenter could only see one. In one condition, the experimenter would ask *'Where's the [novel word]'*, whereas in the other she would say *'There's the [novel word]'*, and then ask the child '*Can I have the [novel word]?*. In both conditions, the experimenter's gaze was fixed on the visible object. While both monolinguals and bilinguals selected the visible object above chance in the 'there' condition, only bilinguals selected the hidden object above chance in the 'where' condition, using the speaker's linguistic information in combination with (or despite of) the gaze cue. Bilinguals also seemed to attend more to speaker's cues compared to monolinguals when confronted to incongruent information for word learning. (Colunga et al., 2012) tested simultaneous balanced bilinguals (aged 24-36 months, English-L2) in a word-learning paradigm pitching object property cues against speaker's cue (eye gaze). In this experiment, children were presented with a novel word and a novel object and asked to find 'another one' among an array of 8 different objects matching the target object in shape only, colour only, texture only, or a combination of these dimensions (i.e., shape and texture, shape and colour, etc.). The objects were grouped in a way that all the shape matches (4) were on one side of the table whereas all the remaining objects (2 matching in colour and texture, one in texture only and one in colour only) were on the other side. In the congruent condition, the experimenter's gaze was directed at the shape matching side, whereas in the incongruent condition it was directed at the other side of the table. Results show that monolinguals chose shape matches significantly more often than bilinguals, and that bilinguals chose the shape matching objects marginally more in the congruent trials, whereas monolinguals showed no differences between the two types of trials. Another recent study by Groba, De Houwer, Mehnert, Rossi, and Obrig (2018) conducted on 3 to 5 year old simultaneous bilinguals (Spanish-German) using a similar paradigm to investigate reliance on object property and pragmatic (gesture) cues in adjective learning yielded no behavioural differences, but found higher activation in bilinguals of an area related to gesture interpretation. Other studies by Yow and colleagues also found bilinguals to rely more on speaker-related than other types of cues for interpretation: emotional affect versus semantic content (Yow, 2015) and eye gaze versus object salience and mutual exclusivity (Yow et al., 2017) compared to monolinguals.

2.3.2 Types of pragmatics and pragmatic inferences

There is a growing body of evidence that pragmatic enrichments do not always require Theory of Mind and that a distinction probably needs to be drawn between those that typically do and those that do not (Kissine, 2012; Sperber, 1994). Hochstein, Bale, and Barner (2018) recently found adolescents diagnosed with ASD (Autism Spectrum Disorder) and neurotypical controls to perform equally well on scalar implicatures and in a task where they were explicitly asked to reason about speaker's knowledge state ('ignorance' implicatures). However, unlike neurotypicals, the ASD group continued to compute the scalar implicature in a condition where it was not warranted by speaker knowledge state. This suggests that these types of implicatures do not normally require epistemic reasoning and that the ability to reason about epistemic states when explicitly asked does not entail the ability to make spontaneous 'online' use of this skill when it *is* required (i.e., for implicature cancellation).

While some authors have started distinguishing between 'linguistic' and 'social' pragmatics (Andrés-Roqueta & Katsos, 2017), the overlap between source of information and use of ToM is, as we have seen, not necessarily perfect: some pragmatic processes triggered by linguistic cues do require ToM reasoning (e.g., irony), whereas non-linguistic social cues such as eye gaze and pointing can be used either directly or to reason about speaker's intentions.

This highlights the need to classify pragmatic processes using multiple, potentially orthogonal, criteria. Triggers or sources of information can be *linguistic* (words, syntax, previous discourse) or *non-linguistic* (eye gaze, pointing, object manipulation), some enrichments involve ToM and reasoning about *speaker's mental state*, some are more *inferential* ('calculated') in nature, while others are more *defaul*t (cf. generalised implicatures, Levinson, 2000) or more akin to simple recognition. Finally, inferential pragmatics can be based either on the *specificity* (i.e., 'narrowness'), or *relevance* (i.e., relation with context/other cues) of a given meaning. Other important dimensions discussed in the next chapters are *emphasis*, or the degree to which attention is directed towards the provision of the cue (e.g., by means of prosodic stress), and *ostension*, or the degree to which the cue appears to be provided with instructive purposes (e.g., in a teaching setting). A cooperative and economical speaker is assumed to match utility and effort (i.e., to be informative). Given a provided cue and level of effort, assuming a more specific or more related meaning will raise utility. On the other hand, emphasis and ostension (which require additional energy/effort) will then raise the assumption of utility for a particular cue[1]. In this work we focus mostly on inferences based on specificity.

[1]This utility is often referred to as relevance, but the former term is ambiguous having both the meanings of ‘relation’ and ‘usefulness’

Chapter 3

Prosody and fast mapping

All cats are mortal. Socrates is mortal. Therefore, Socrates is a cat.

— Eugene Ionesco

3.1 Pragmatic inference and alternatives

The need for pragmatic inference derives from the *ambiguity* which characterises natural language, i.e., the presence of multiple potential meanings, or mappings between linguistic expressions and referents. One way of assessing each meaning is to assign probabilities to them according to the *specificity* of each hypothesis: all things being equal, more informative or specific hypotheses will be preferred (the *Size Principle*, Tenenbaum & Griffiths, 2001). In addition, like all living organisms, humans are lazy. Linguistic communication is no exception to this rule and hearers can generally assume cost/benefit efficiency in production to infer speaker meaning. For example, if a professor utters *'Some students failed the test'* after marking, the hearer can infer that not all students failed by assuming that, at equal 'price' (i.e., linguistic cost) said professor could have used the more specific *'all'*. This is an example of *scalar*

Fig. 3.1 Contrastive inference (ad hoc implicature). Stiller et al., 2015

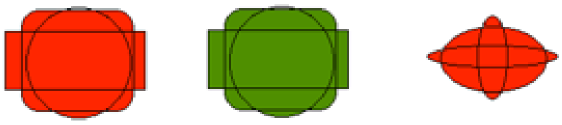


Fig. 3.2 Contrastive inference

implicature, a name which refers to the scale of semantic strength it is said to derive from (*none-some-all*, etc. L. R. Horn & Ward, 2004). Following Grice's (1975) first Maxim of Quantity, the implicature is derived by assuming that the speaker is, at equal cost, saying 'as much as possible'. The maxim can also be used on a scale that is not semantically formalised but has been generated from context to derive an *ad hoc* implicature. For example, if presented with two characters, one with hat only and the other with both items (see Figure 3.1), 3-4 year old children and adults asked to pick the referent with prompt *'My friend has glasses'* reliably choose the character with glasses only, inferring that the speaker would have used the more specific feature *'hat'* if referring to the character with both items, in a process similar to the classic 'some but not all' scalar implicature)(Stiller et al., 2015).

Contrastive inference is another type of implicature using Grice's second Maxim of Quantity: 'do not say more than you need'. Hearers presented with an array of four objects and prompted with an utterance beginning with *'Pick up the tall...'* will start shifting their eye-gaze towards the tall glass (which has a small glass counterpart) rather than the tall jug before hearing the noun, inferring that the modifier is more specific, being required, for this referent (Nadig & Sedivy, 2002). Similarly, *'the red dax'* should be taken to refer to the leftmost rather than the rightmost object in Figure 3.2. In all examples outlined above, the given expression is more specific to the target referent by virtue of being the only possible choice for this referent ('some not all' can only be expressed by *'some'*, the character with glasses by *'glasses'* and the leftmost object by *'red dax'*) whereas the alternative referent not only has an available alternative expression sufficient for identification (*'jug'*), but sometimes even a more informative one (*'all', 'hat'*).

Both types of quantity implicatures thus have in common that they appear to be the result of dismissing other meanings or referents as having a more appropriate available description than the one that was used (The *Alternatives Hypothesis*, Barner & Bachrach, 2010). As hearers rapidly assess each potential referent in the relevant array (or set of semantically compatible meanings), salient alternative ways of referring to each candidate will come to mind and, if considered more informatively efficient, lower the candidate's probability to be chosen as the intended target. Which candidate is assessed first is a matter of debate between two theoretical views: the two step *literal model* (where the logical meaning is assessed first, Huang & Snedeker, 2009) and the *generalised/default implicature hypothesis* (where the implicature is generated by default and subsequently cancelled if not warranted by context, Levinson, 2000). In the scalar/*ad hoc* implicature case, the expressions *'some'* and *'glasses'* would be underinformative to refer to 'all' and to the character with hat and glasses, whereas in the contrastive case the use of the modifier *'tall'* for the jug would be considered overinformative.

In the RSA (Rational Speech Act) framework (Frank & Goodman, 2012), which models pragmatic inference as Bayesian inference, this translates into quantitative predictions of meaning assignment for a given referent r directly proportional to the relative specificity of the target hypothesis or mapping with the linguistic expression e (if we assume a uniform prior probability for all referents, i.e., that all referents have a priori the same probability of being chosen/are equally salient, otherwise this likelihood is differentially weighted by the prior to yield posterior probability). Thus, for each potential referent/hypothesis we have

$$P(r|e) \propto P(r)P(e|r) \tag{3.1}$$

Where P(r|e) is the posterior probability of a referent r given an expression e, which is proportional to the prior probability of the referent being referred to (in the absence of any expression) P(r) multiplied by the likelihood P(e|r) of an expression being uttered given that the speaker has a referent r in mind. Given that for both types of inferences described above (scalar/*ad hoc* and contrastive) the target hypothesis is twice as specific (i.e., the expression is used with half as many referents) as the alternative hypothesis, it would be predicted to be chosen twice as often, i.e., 0.67 vs. 0.33.

3.2 Contrast, prosody and development

In a seminal study of contrastive inference, Gelman and Markman (1985) presented 3 and 4 year old children with a display of four objects: a modified target object among three of its kind (e.g., a red butterfly next to a blue and a yellow butterfly) and a modified single competitor (e.g., a red chair) and instructed them to select an object with a modified noun phrase involving a pronoun (e.g., *'Show me the red one'*). While 4 year olds reliably chose the object with counterparts (e.g., the red butterfly) by inferring that 'the red one' would be overinformative for the single competitor (the

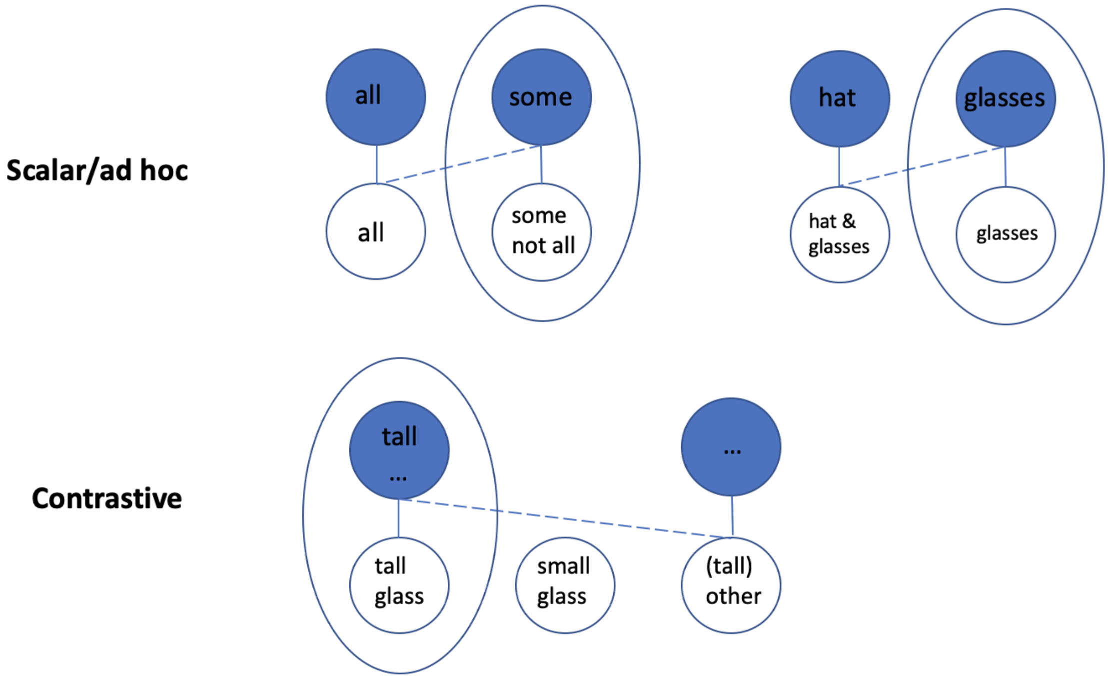


Fig. 3.3 Scalar and contrastive inferences *(linguistic expressions in blue, referents in white, dotted lines represent alternative meaning hypotheses)*

red chair), 3 year olds performed at chance. This is in contrast with the results of Kronmüller et al. (2014). In their study, Kronmüller et al. (2014) presented seven-year olds, ten year-olds and adults with a paradigm whereby participants observed an exchange between experimenter and listener. The exchange was about two unknown cards held up by the experimenter. The cards were picked from a set of four cards which were visible to the participant (i.e., the participant knew that the two unknown cards were picked from a set of, e.g., a brown rabbit, a white rabbit, a goat and a duck). The participant heard the experimenter ask the listener to point at one of the two unknown cards they were holding up by referring to it using either a modified or unmodified noun phrase (e.g., *'Show me the brown rabbit* or *'Show me the rabbit'*). The participant was then asked *'What do you think is on the other card?'*. Adults made the expected contrastive inference both in the modified condition (e.g., if the speaker used 'the brown rabbit' it means that the other card is the white rabbit) and the reverse exclusive inference in the unmodified condition (e.g., if they used 'the rabbit' it means that the other card is NOT the white rabbit). Ten year olds, on the other hand, only succeeded in making the inference in the modified condition, and at a much lower rate than adults (54% vs. 78%). Seven year olds did not appear to make any inference at all and performed at chance in both the modified and non-modified conditions.

The discrepancy between Gelman and Markman (1985) and Kronmüller et al. (2014)'s results could perhaps be explained by the more complex nature of the task in Kronmüller et al. (2014), which involved a metalinguistic component (*'What do you THINK is on the other card?'*) and an indirect question meaning the action taken was not directly a result of the noun phrase used. Moreover, unlike in Gelman and Markman (1985) where the choices were performed directly on the relevant set, Kronmüller et al. (2014)'s paradigm involved having to *visually project* the content of the cards displayed in front of the participant into the unseen ones in the experimenter's hands. Even though the contents were continuously visible to the participant in their own set of cards, there was still a processing effort required to 'make a copy' of them for the unseen cards and to reason by going back-and-forth between the visible and invisible cards. As the authors noticed themselves, this could cause confusion, for example regarding which set or subset was being taken into account by the speaker (e.g., if the speaker was taking the whole set into account, the brown rabbit would have had to be referred to as 'the brown rabbit' no matter what card was held up with it). While the authors corrected this particular confusion by making it clear in a subsequent experiment that only the participant had access to the full set, it is likely that the projection and back-and-forth reasoning effort might have proven particularly difficult for young children with limited processing abilities. However, other evidence indicates that preschoolers do lag behind adults in terms of the strength and timing of contrastive inferences: in an online eye-tracking study conducted by Sekerina and Trueswell (2012) with Russian speaking 4 to 6 year olds who were presented with a display similar to Sedivy, Tanenhaus, Chambers, and Carlson (1999)'s, children did show facilitation of processing in the expected contexts (e.g., a Russian split construction which highlights a contrastive reading or a display where only one modified object has a counterpart) but only *after* hearing the head noun, unlike adults who displayed anticipatory looks towards the target as soon as they heard the adjective.

While structural constructions such as focus or modifier are one way of creating a contrastive inference, another way involves the use of prosody and, in particular, lexical stress (also referred to as emphasis)[1]. For example, while a hearer might not derive any particular implicit meaning from the utterance *'It's a very nice garden'*, they could infer a contrast if prosodic stress is used to emphasise a particular word (*'It's a very nice GARDEN' e.g., [as opposed to the house next to it]*). The interpretation and realisation of this prosodic stress appears highly dependent on context: the same noun-final stress (*'Give me the yellow ROSES'*) can take either a neutral or contrastive reading and contrastive stress could be realised with either a simple H* target high or fall-rise L+H* (according to autosegmental theories, cf. Pierrehumbert & Hirschberg, 1990). However, when confronted with the emphasised noun (*'It's a very nice GARDEN'*), adults reliably prefer a picture showing a nice garden and house falling down over one with both a nice garden and house. 10 year olds, on the other hand, do not show such a preference (Cruttenden, 1985). While Sedivy et al. (1999, Experiment 2) failed to find a significant facilitative effect of prosodic stress on modifier for contrastive inference, they did find signs of it in previous work (Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard, & Carlson, 1995) and suggested that its absence in their later study might have been due to ceiling effects, as some fillers did not include adjectives (drawing attention to their presence) and the use of adjectives in a first preparatory instruction might have increased attention towards the contrasted target object in the critical instruction.

[1]From now on we use capital letters in utterance examples to signify the use of prosodic stress.

Indeed, other subsequent studies did find prosodic stress to positively impact early looks towards target objects in adults when hearing the adjective (Sekerina & Trueswell, 2012; Weber, Braun, & Crocker, 2006) and to have the same effect in children after the noun had been heard (Sekerina & Trueswell, 2012).

Despite young children's well known general sensitivity to prosodic variation and their preference for infant-directed speech(Fernald, 1985), what Cruttenden (1985) and other similar studies seem to show is that prosody alone is not a reliable or salient enough cue for children to derive a contrastive inference, but could facilitate such inferences if presented with other converging or supporting cues. This is what is suggested by other studies which found that preschoolers manage to perform contrastive inferences which they previously failed at when given additional discourse support for a contrastive reading (Horowtiz & Frank, 2014 Experiment 1) or explicit access to alternatives in the previous discourse (Kurumada & Clark, 2017; Horowitz & Frank, 2014, Experiment 3).

Why did the preschoolers in these studies fail in first instance at deriving contrastive inferences while they succeeded in Gelman and Markman (1985)? In Horowitz and Frank (2014), children were given a description and picture of a specific individual object from a novel category (e.g., *'This is a broken tibu'*) and had to extend the novel category (e.g.,*'What do you think tibus usually look like?'*) by choosing between an unmodified option (e.g., an unbroken tibu) or an option modified as the training example (e.g., another broken tibu). Besides containing (again) a metalinguistic component, this task involves creating a mental representation of the novel category from one exposure, but also *inhibiting* a salient response to select an object which is similar or identical to the training object. Similarly, in Clark and Kurumada (2017), four year olds heard an utterance which involved prosodic stress inducing a contrastive interpretation (e.g., *'It LOOKS like a zebra'* [but it isn't]) or did not (e.g., *'It looks like* *a zebra'* [and it is]) and had to select between a picture of the mentioned animal or a (correct) contrastive picture depicting another animal (e.g., a picture of a zebra or a picture of an okapi). Here again, participants had to not only recognise prosodic stress and infer the alternative expression, which was not explicit (i.e., 'the speaker could have said *'It is a zebra'* if they thought that it was'), but also to inhibit a response based on matching noun and picture (i.e., selecting the zebra when they heard the word 'zebra'). Probably due to these additional processing demands, children in both studies succeeded in making the inference only when provided with further cues to support it: in Clark and Kurumada (2017), this was achieved by explicitly providing an alternative expression (alternating between utterances of the type *'It is a zebra'* and *'It LOOKS like a zebra'*, thus providing utterance forms which differed in two ways and drew children's attention to the contrast. In Horowitz and Frank (2014) better children's performance came from either making the modifier more salient (e.g., *'This is a special kind of tibu.This is a broken tibu'*) or from explicitly providing alternative modifiers in the previous discourse (i.e., reading a 'book of opposites' with the child where objects were described in contrasting pairs, e.g., 'dirty'/'clean', 'wet'/'dry', etc.)

Contrastive stress serves the purpose of drawing attention to a particular component of discourse, leaving the hearer to figure out the contextual relevance of the additional energy expended to emphasise this component. By exploiting an asymmetry in energy and resulting attention, it is in a sense one-directional (either the speaker drew attention to this component or they did not). Emotional affect, on the other hand, is a prosodic cue which does not require the hearer to compute a contextually relevant contrast set, as there are (mainly) two possible options. Since it is (at least) bi-directional, the affect's meaning cannot be arrived at through the same type of reasoning used to interpret contrastive stress. In other words, while the relation 'energy>drawing attention>importance' can be inferred, the association between the particular prosodic characteristics of a sad or happy (or fearful) tone and the underlying emotions they represent have to be learned in order to be recognised, just like we learn to associate colour words with the range of wave lengths that each denotes. As Quam and Swingley (2012) point out, children's early sensitivity to pitch patterns does not entail an ability to correctly understand or *interpret* these patterns. While 12 month old infants do show sensitivity to a fearful sounding voice compared to a neutral one (Mumme & Fernald, 2003), this is again a case of one-directional inference. Infants have been shown to generally pay little attention to paralinguistic cues: they do not improve recognition of face expressions in 5 month olds (D'Entremont & Muir, 1999) and children, unlike adults, rely on lexical over prosodic information for interpretation until as old as 10 years old (Morton & Trehub, 2001). However, evidence shows that 5 year olds are at least able to use negative affect to correctly infer the extension of a novel word given a damaged versus enhanced version of a previously seen novel object (Berman, Graham, Callaway, & Chambers, 2013).

3.3 Word learning and social cues

The socio-pragmatic theory of word learning holds that the process of word learning is 'inherently and thoroughly social' (Bruner, 1985; Tomasello, 2000). As such, it has traditionally encompassed as 'socio-pragmatic' cues everything in the context of the utterance that could be used to retrieve the intended referent for a label, including speaker-related and social information, prior and immediate, linguistic and non-linguistic. Immediate socio-pragmatic information comprises everything that is necessary to achieve precisely targeted joint attention, both linguistic (carrier sentences) and paralinguistic (gestures, stance, facial expressions, prosody, actions). Prior pragmatic information constitutes the common ground between speaker and hearer, determining givenness or novelty levels of referents, speakers and labels involved, again both through linguistic (anaphoras) and non-linguistic (familiarity with potential referents) means. Young children have time and again demonstrated efficient and early capacity to use these social cues, learning words using eye gaze (Baldwin, 1993), gestures (Horst & Samuelson, 2008) ; facial expressions (Tomasello, Strosberg & Akhtar, 1996), adult intentional/accidental (i.e., clumsy) behaviour (Carpenter, Akhtar, & Tomasello, 1998) ; context in which the referent is presented (i.e., other objects/actions that accompany it) (Tomasello & Akhtar, 1995; Waxman & Booth, 2001), novelty/givenness status (Moll & Tomasello, 2007) and prosody (Grassmann & Tomasello, 2010).

However, talking about a 'socio-pragmatic' account of word learning might be conflating two distinct sets of skills: on one hand the capacity to attend to these types of social cues, on the other a separate competence in making 'true' pragmatic inferences.

According to Grice (1975), communication proceeds successfully by assuming the speaker is cooperative, i.e. behaves according to a set of maxims enjoining her to be truthful, relevant, clear and informative. Building on these assumptions, listeners can try and retrieve the intended meaning from the speaker's utterance or communicative behaviour. In the case of word learning, the purpose of the learner is to find which referent is attached to the speaker's label. Using the cues mentioned above, they could possibly be doing so by computing a pragmatic inference (i.e., an inference that uses hypotheses about the speaker's communicative intentions to understand the meaning of his words or actions): *if the speaker is gazing or pointing at this object, it is probably because they intend to refer to it*; *if they are using a novel label, it is probably because they have a new referent in mind* or *if they are pairing this pink elephant with other animals, they probably mean the label to refer to a category; if they are pairing a pink elephant with other pink objects, they probably intend to refer to the property*, etc.

Alternatively, in many of these cases it could be argued that they reach the same conclusion by using these cues in a more direct manner, i.e. because they contribute to raising the salience of one referent over the others, without any intention-reading actually having taken place (Ambridge & Lieven, 2011). This is what is suggested by Frank (2014) and Frank et al. (2009)'s model, which separates two dimensions for word learning processes: the source of the cue (cross-situational or social) and the way the cue is used (in a purely associative manner or through the computation of intentions). Thus the social cues could contribute to contrasting the target referent with other potential referents directly by highlighting it, or the contrast could be the result of a pragmatic inference about what the speaker intended to highlight. Of course using both sources of information and contrast will be most efficient. Indeed, cross-situational and social information used directly will only contribute differentially to the contextual salience of each particular referent, whereas using speaker's intentions will allow them to gauge the relative informativeness of the speaker's utterance with respect to each of the referents.

As we have seen, contrastive inference is a type of pragmatic inference where the reason for providing a certain cue (e.g., an adjectival modifier, or prosodic stress) is assumed to be for informative purposes (i.e., to distinguish one potential referent from another).

Emotional affect (expressed via positive or negative intonation), on the other hand, is a type of socio-pragmatic cue which, like eye gaze and pointing, could potentially be used directly to identify a particular referent with a positive or negative feature without the need to reason about *why* the speaker provided the cue. In this scenario, the affect (or point, or gaze) would shift the child's attention towards the target object, raising its probability as a candidate for the novel word without any processes involving Theory of Mind taking place (Frank, 2014). This can to a certain extent be related to

Grice (1975)'s dichotomy between natural meanings (which can be used directly) and non-natural meanings (which require the understanding of a communicative intention).

We have seen that children are able to use social cues for word learning from a very young age (around 10-12 months). Yet they have been shown to have trouble with computing some types of pragmatic inferences before 7 years old (Barner, Brooks, & Bale, 2011; Katsos, 2009), even though they appear to be able to make contrastive inferences similar to the ones described above around 3-4 years old (Gelman & Markman, 1985; Frank & Goodman, 2014). However, in some studies even these types of inferences are not seen before 5- or even 10 years old (Kronmüller et al., 2014; Morisseau, Davies, & Matthews, 2013)). It is possible that infant's first use of social cues (and possibly at least part of their subsequent use) relies on simple associative mechanisms and salience and allocating attention efficiently and appropriately (i.e., focussing on speaker-related information) without requiring intention-reading through Theory of Mind, or the actual computation of a pragmatic inference. This is not to say that children always use social cues as attentional devices and do not engage in inferences about intentions (studies have repeatedly shown that infants are able to find the referent of a novel word as the one the speaker has in mind but is not physically present or accessible, e.g., Akhtar, Carpenter, & Tomasello, 1996; Tomasello et al., 1996), but simply that they do not necessarily do so in each case. As Ambridge and Lieven (2011) point out, while there have been many studies showing that infants are able to understand other's intentions, only a few have actually demonstrated that they use this understanding to choose between potential referents for a word (Diesendruck, 2005; Diesendruck & Markson, 2001; Tomasello & Akhtar, 1995). It is possible that in certain contexts, pragmatic understanding which has traditionally been seen as the result of intention-reading is in fact obtained through more general mechanisms using the linguistic information or social cues in a more direct way and 'shortcutting' the pragmatic process. The existence of such an 'egocentric' pragmatic competence would then contribute to explain some pragmatic understanding displayed by children with autism spectrum disorders (Kissine, 2012; Kissine et al., 2015). If this is the case, these two sets of skills potentially need to be distinguished as requiring different types of abilities (differential allocation of attention/ToM and inferencing) and having separate developmental timelines. Moreover, using social cues (or speaker-related information) should not be confused with exerting a pragmatic competence (or making inferences about speakers' intentions).

Given the emergent and growing evidence that bilinguals are more competent with pragmatics than their monolingual peers, it is natural to ask at this point what underlies the differences. We discern three possibilities which we outlined in Chapter 2. First, because most of these paradigms used with younger children involved inhibiting an irrelevant cue in order to focus on the socio-pragmatic cue, the results in these studies might have been driven by enhanced general cognitive abilities, such as inhibitory control. There has been suggestions that the constant switching and inhibition needed to monitor multiple languages could lead to enhanced executive abilities in bilinguals (Green, 1998). As we mentioned in Chapter 2, many studies have shown early bilingual children outperforming monolinguals on tasks related to executive functioning skills, such as working memory, inhibition and cognitive flexibility (e.g., Barac & Bialystok, 2011; Bialystok & Craik, 2010). These improved skills in executive control could enhance the basic associative mechanisms underlying instances of word learning, both by retention of a greater number of candidate referents for a novel word, or by allowing easier switching between hypothesised referents. This advantage could be mediated by verbal short-term memory or by inhibitory control (Papagno & Vallar, 1995; Bartolotti et al., 2011). In addition, these cognitive advantages could also have an impact on word learning beyond association by improving their ability to combine cues from different sources and ignore irrelevant information (Yow & Markman, 2015; Kaushanskaya & Marian, 2007). Yow & Markman (2015)'s study suggests that bilinguals could indeed have an advantage for ignoring irrelevant cues in general, not only in favour of a social cue. They presented 3-year old bilingual and monolingual children with a word-learning task where they saw two objects while the experimenter could only see one. In one condition, the experimenter asked *'Where's the [novel word]'*, whereas in the other she would say *'There's the [novel word]'*, and then ask the child *'Can I have the [novel word]?'*. In both conditions, the experimenter's gaze was fixed on the visible object. While both monolinguals and bilinguals selected the visible object above chance in the 'there' condition, only bilinguals selected the hidden object above chance in the 'where' condition, using the speaker's semantic information in combination with (or despite of) the gaze cue.

However, while an advantage in pragmatics could in principle be related to enhanced executive functions, there are two main reasons to doubt that this is the case. At a general level, many studies have failed to replicate the bilingual advantages in cognition and the very existence of this alleged advantage has recently been questioned (Paap et al., 2015). The inconsistency in the results could be due to different types of bilingualism having differential influences: earlier acquisition, higher proficiency and more frequent switching between languages are for example likely to result in a higher need for control abilities (e.g., Luk et al., 2011). However, more pertinent to the literature that we are reviewing here, those studies which measured executive control failed to find a difference between bilinguals and monolinguals which would have explained the former's higher performance in pragmatics (Yow & Markman, 2015;Yow et al., 2017; Champoux-Larsson & Dylman, 2018) or failed to find such a difference for bilingually exposed children despite both bilingually exposed and bilinguals outperforming monolinguals (Fan et al., 2015). Therefore, in line with Fan et al. (2015) we suggest that the bilingual advantage in pragmatics is not likely to be due to enhanced executive functions.

The second possibility we consider is that bilingualism increases attention to speaker or social cues (or preferential weighting of them compared to other types of cues). Young children demonstrate efficient and early capacity to use social cues for word learning (e.g., Baldwin, 1993; Tomasello et al., 1996). As we have seen, according to standard pragmatic theory (e.g. Grice, 1975), communication proceeds successfully by assuming the speaker is cooperative, i.e. behaves according to a set of maxims enjoining her to be truthful, relevant, clear and informative. Building on these assumptions, listeners can try and retrieve the intended meaning from the speaker's utterance or communicative behaviour. In the case of word learning, the purpose of the learner is to find which referent is attached to the speaker's label. Using the cues mentioned above, they could do it by computing a pragmatic inference by reasoning about speaker's intentions. The other possibility is that they reach this conclusion by using these cues in a more direct manner, i.e. because they contribute to raising the salience of one referent over the others, without any intention-reading actually having taken place (Ambridge & Lieven, 2011; Frank, 2014). Therefore, the second possibility that we put forward is that bilingual children, because of the nature of their interactions with their social environment, are better at paying attention to social cues.

Finally, the third possibility is that bilingual experience could contribute to enhance pragmatic competence per se, i.e., the ability to reason about communicative intentions and informativeness. This is a distinct possibility from the one mentioned above. True pragmatic inference distinguishes itself from other types of inference in that it requires not only recognition of the cue provided but also reasoning about the speaker's intentions in providing the cue. Take for example contrastive inference, a type of implicature using Grice's second Maxim of Quantity: *'do not say more than you need'*.

In this type of pragmatic inference, the reason for providing a certain cue (e.g., an adjectival modifier, or prosodic stress) is assumed to be for informative purposes (i.e., to distinguish one potential referent from another). As we have seen, hearers presented with an array of four objects and prompted with an utterance beginning with *'Pick up the tall...'* will start shifting their eye-gaze towards the tall glass (which has a small glass counterpart) rather than the tall jug (which doesn't have another jug as a contrast) before hearing the noun, inferring that the modifier is more specific, being required, for this referent (Nadig & Sedivy, 2002). Emotional affect, in contrast, can potentially be used directly to identify a particular referent with a positive or negative feature without the need to reason about speaker's intentions, simply by raising the salience of one candidate over the others. This difference, namely between the mere recognition of a cue and the actual use of a cue into a chain of pragmatic inferential reasoning, is related to Grice's (1975) dichotomy between natural meanings (which can be used directly) and non-natural meanings (which require the understanding of a communicative intention). Therefore, the third possibility for bilingual children's better performance with pragmatics could be that the nature of their linguistic experiences have made them more adept at pragmatic reasoning per se.

To take stock, better performance by bilingual children in the social cues word learning studies previously mentioned could be the result of a bilingual advantage in general cognitive abilities (and inhibition of irrelevant information) or to increased attention towards speaker cues, or they could result from increased ability to reason about speaker's intentions in providing the cue. While there is growing evidence that suggests the first possibility is not promising, the other two are very much up for empirical scrutiny. The current study sought to investigate these last two hypotheses by testing monolingual and bilingually exposed children's performance on two different word learning/fast-mapping paradigms in the absence of conflicting information (i.e., where success could not be obtained by ignoring an irrelevant cue). These are a contrastive inference task (which required reasoning about informativeness) and an emotional affect task (which could be completed by simply attending to speaker cues). We ask two interrelated questions. First, if bilinguals perform better than monolinguals in these two tasks where attending to speaker cues does not require ignoring another type of cue (which is unlike the tasks used in the bilingual word-learning literature to date). Critically, we ask if bilinguals perform better in a word learning task where they are required not only to attend to a speaker cue but also to reason about it (contrastive stress) compared to a task that requires mere attention to a speaker cue (emotional affect).

3.4 Experiment 1: contrastive inference

The purpose of this experiment was to investigate whether bilingual children would outperform monolingual children in a fast mapping task where they had to pick the referent of a novel word on the basis of reasoning about the communicative intention behind a cue given by a speaker. In the experiment, children heard a prompt containing a novel word which had been modified either with a non-stressed or a stressed adjective (e.g., *'Touch the WET/wet gorp'*). They were shown a display with four novel aliens, two of which were semantically compatible with the prompt (ie., two of the aliens were wet). However, only one of the compatible aliens had a dry counterpart which made the use of a contrastive modifier more likely. Given our hypothesis that the previously evidenced socio-pragmatic advantage in bilingual children is not uniquely the result of better inhibitory control or increased attention to speaker cues, we predict that bilingual children would show a better ability to use the contrastive inference to identify the referent of the novel word despite the fact that the modifier itself does not raise the salience of either compatible referent above the other and thus inhibitory control or better attention to the modifier would not lead to better performance.

The paradigm we used was similar to Gelman and Markman (1985) with the difference that these authors used the word 'one' and familiar objects (in Study 1) and novel adjectives and objects (in Study 2) (e.g., *'the red one'* or *'the fep one'*) instead of novel nouns as we did. In addition, we added a condition where the modifier was prosodically stressed to investigate whether performance would differ when this additional cue was available.

3.4.1 Method

Participants

The study was approved by the Cambridge Psychology Research Ethics Committee and parents in participating schools were sent information about the study along with a form allowing them to opt-in or opt-out depending on the school's policy. In total, 270 children aged 4-6 years old were recruited through schools in Cambridge and London. Demographic and language information was obtained through parental forms for 74 children, while the information for the remaining participants was obtained through school staff and proved to be highly reliable (over 98 % match with the 74 questionnaires) when compared to available parental information. Of the 270 children, 138 (66 females, 72 males, mean age=5;2, sd=6.9 months) had been exposed daily to a second language for at least a year, including children in French immersion program (n=26), children identified through parental forms as sequential bilinguals (n=18) or as simultaneous bilinguals (n=26) and identified as bilingual by their teachers (for these children the school described the child as having English as additional language (EAL; n=68) which is a technical term used in the UK educational setting to signify children whose dominant language is not English). The remaining 132 children (62 females, 70 males, mean age=5;5, sd=8.2 months) had had no regular exposure to a language other than English. The average percentage of free school meals (FSM) in each of the participating schools was used as a proxy for socio-economic status (see Hobbs & Vignoles, 2007, for a discussion on the use of FSM for this purpose). Languages other than French (immersion and home, n=34) were Hindi (n=15), Gujarati (n=13), Tamil (n=10), Romanian (n=10), Urdu (n=6), Polish (n=6), Arabic (n=5), Pakhto (n=3), Portuguese (n=3), German (n=3), Cantonese (n=2), Czech (n=2), Somali (n=2), Slovak (n=2), ESL (n=2), Albanian, Bahasa, Farsi, Greek, Hungarian, Italian, Lithuanian, Malayalam, Mandarin, Punjabi, Persian, Russian, Serbian, Sindhi, Sinhala, Swahili, Swedish, Thai, Turkish and Vietnamese (all n=1).

75 monolingual adult participants were recruited through Prolific Academic for piloting purposes and to uncover any developmental effects. Adult participants completed the study online for a reward of £0.87.

Receptive vocabulary task

Children first completed a computerised version of the BPVS-3 (Dunn & Dunn, 2009) implemented on the touchscreen laptop to test receptive vocabulary. This is a picture-matching task in which children are asked to point to one of four pictures that matches the word uttered. The items are arranged in blocks of 12, and the test continues until children have made 8 or more mistakes in a block. Children received two warm-up trials. Raw scores rather than standardised were used in the analyses as they indicated both vocabulary and developmental levels. Receptive vocabulary scores are also a more direct measure of the impact of socioeconomic factors on language development (the issue of interest here) than SES indices such as maternal education or household income. Instructions were recorded by a female native English speaker.

Stimuli

Stimuli were 12 pictures each with four unknown aliens, two of which were of the same kind (e.g., two type A aliens, one type B alien, one type C alien). Pictures for the critical conditions contained a target and distractor featuring the target property (e.g., a wet type A alien and a wet type B alien) and a counterpart which did not feature the property (e.g., a dry type A alien). Pictures for the control condition contained only one alien featuring the target property. The properties were chosen so that they could be described by common adjectives, would be accidental/non-intrinsic (so that one alien could feature the property while its counterpart did not) and could be used to describe modified or unmodified creatures (e.g., a wet alien or dry alien) to account for the potentially increased salience of modified aliens. Colours were avoided because of the confusion often displayed by children around that age in reliably recognising and naming them (Bornstein, 1985). An example stimulus for the stressed and non-stressed conditions can be found in figure 3.4, and for the control condition in figure 3.5. A full list of stimuli (utterances and pictures) can be found in Appendix 1.

Procedure

For adult participants stimuli were presented using the Qualtrics platform. For children participants stimuli were implemented and presented on a touchscreen laptop using Superlab version 5.0 (Cedrus, San Pedro, CA). Clicks were recorded as raw X and Y pixel coordinates of the point where the screen had been touched and answers were subsequently coded by matching the coordinates to the corresponding answer area among the four possible choices on the picture. Children were tested in a quiet room at their school. They were introduced through the computer to Mr. Puppet, who had recently made some alien friends and lent them some toys which he now needed to get back. They were then asked if they would like to help Mr. Puppet find the aliens, and

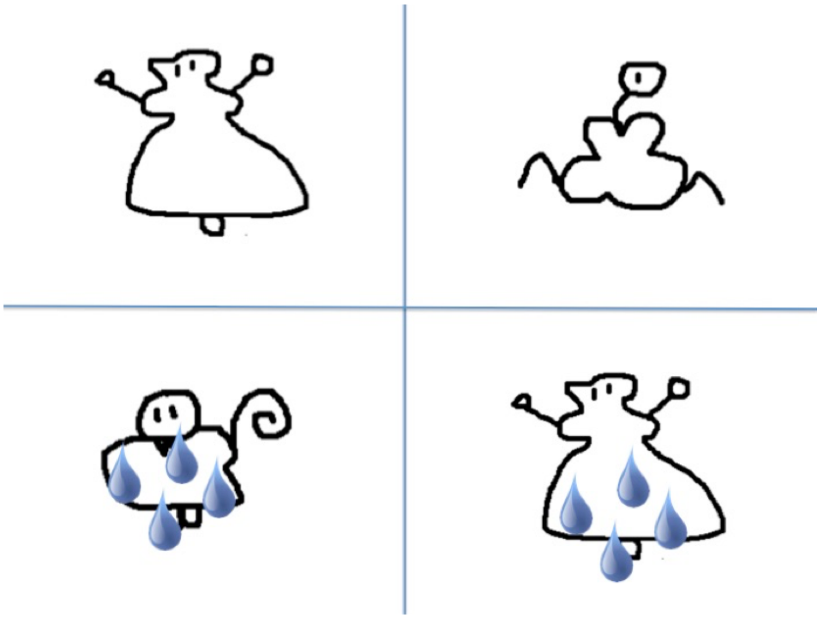


Fig. 3.4 Example stimulus experiment 1 (non-stressed and stressed conditions) *'Touch the wet gorp/Touch the WET gorp'.*

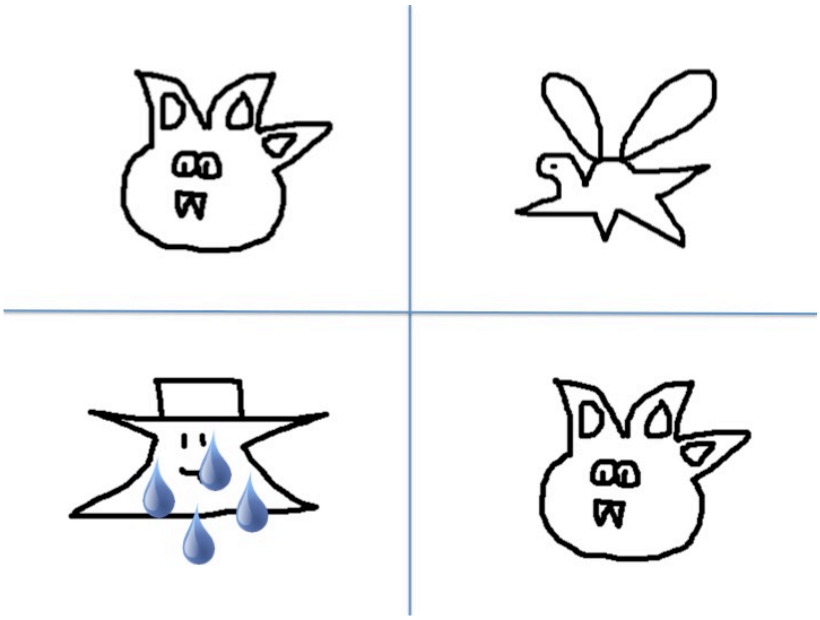


Fig. 3.5 Example stimulus experiment 1 (control condition) *'Touch the wet gorp'.*

given a test trial where they had to find the 'wet gloop' and 'dirty gloop' (feedback provided) before proceeding to the task. There were four trials of each type (12 total): non-stressed condition (e.g., *'Touch the wet gorp'*), stressed condition (*'Touch the WET gorp'*) and control (similar instructions to the non-stressed condition but only one alien had the target feature).When the participant had made their choice by touching the screen, the next trial appeared. Each trial was preceded by a fixation cross centered on the screen. Instructions were recorded by a female native English speaker using a Sennheiser ME 64 cardioid microphone connected to a Tascam HD-P2 Compact Flash Audio Recorder. Recordings were made in 24bit mono with a sample rate of 44.1kHz . Experimental design was within-subject, with 4 target features/adjectives ('wet', 'dry', 'clean' and 'dirty') and 12 novel words (*gorp, pitack, rapook, lep, plonk, yubba, moozie, ral, flurg, dinkoo, patam* and *tweep*). There were two lists of items counterbalanced for word/alien pairings and target position. Trials were randomised.

3.4.2 Results

Final results can be seen in Table 1. Data from eleven monolingual children were excluded from final analyses because of experimenter's error or child fussiness or failing to complete the task. We first conducted analyses with target and distractor responses only (removing errors where the character selected did not semantically match the instruction, e.g., children chose a dry character in response to a prompt for a 'wet gorp') to be able to set chance performance at 0.5 and examine each group's performance in each condition (chance with four choices would otherwise technically have been set at 0.25). In the regression model raw scores (including all types of errors) were used to allow comparison with performance in control condition. We first conducted t-tests on answers excluding non-distractor errors to evaluate performance against chance in each group. This first set of analyses showed that adults were significantly above chance in selecting target in both stressed (m=0.66, sd=0.47, t = 3.19, df = 350, p= 0.001) non-stressed condition (m=0.58, sd=0.49, t = 6.35, df = 345, p<0.001) and control (m=0.92, sd=0.35, t=2.18, df=350, p<0.001). Children, on the other hand, were significantly above chance in the stressed condition (m=0.56, sd=0.49, t = 4.06, df = 947, p<0.0001) but not in the non-stressed condition (m=0.51, sd=0.50, t = 0.59, df = 927, p= 0.55). This pattern applied to both monolinguals (non stressed=0.50, p=0.3461; stressed=0.58,t = 3.66, df = 489, p=0.0003) and bilinguals (non stressed=0.52, p=0.1543; stressed=0.55, t = 2.06, df = 457, p=0.04).

Preliminary analyses revealed that, compared to the monolinguals, the bilingual group was on average significantly younger (5;1 vs. 5;5, t = 2.63, df = 364.9, p= 0.009), had significantly lower English vocabulary (British Picture Vocabulary raw scores 76 vs. 68,t = 6.3, df = 377.75, p <0.0001) and significantly lower SES (as calculated through averaging free school meal percentages previously normalised using national average and standard deviation,-3.00 vs. -2.51, t = -3.95, df = 372.32, p <0.0001). These were then entered as covariates in our regression model. As all these variables, chronological age, SES and general vocabulary proficiency in the language of testing are known to independently have an effect on performance on linguistic tasks, they were then entered with gender (also shown to have an effect on pragmatic performance, cf. Stiller et al., 2015) as a logistic regression mixed model (D. Bates, Mächler, Bolker, & Walker, 2014) with items and participants as random effects and starting with all design-relevant fixed effects as random slopes, i.e., a maximal random effects model. Following a procedure used in cases of over-parameterisation (Cane, Ferguson, & Apperly, 2017; Horowitz, Schneider, & Frank, 2018), random effects were removed only when they led to non-convergence. Because age and raw BPVS scores were highly correlated, only BPVS scores were retained as they proved to be a better predictor of performance. The final model included gender, normalised BPVS scores, normalised FSM percentages, bilingual status, condition and their interaction as fixed effects, and subject intercepts as random effects. The control condition was used as the reference level.

Performance was significantly correlated with higher vocabulary scores (est=0.23, se=0.04, z=5.45, p <0.0001), lower FSM percentage (est=-0.16, se=0.04, z=-4.35, p <0.0001), bilingual status (est=-0.71, se=0.21, z=-3.39, p=0.0007) and gender (est=-0.22, se=0.09, z=-2.46, p=0.01), with monolinguals and females performing better. Performance was significantly lower in both stressed (est=-2.26, se=0.19, z=-11.93, p <0.0001) and non stressed condition (est=-2.62, se=0.19, z=-13.83, p<0.0001) compared to control. In addition, there was a significant interaction between condition and bilingual status, with bilinguals performing significantly better than expected from their control performance compared to monolinguals in both stressed (est=0.49552, se=0.24059, z=2.060, p=0.04) and non stressed condition (est=0.70183, se=0.24063, z=2.917, p=0.003). A graph showing the results for the adults can be seen in figure 3.6, and for children in figure 3.7. The results broken down for bilingual and monolingual groups can be seen in figure 3.8.

We conducted further analyses on a subset of children including only those who had a maximum score on the control task. However, as this meant the removal of data from over a hundred children, and given that the contrastive inference effect had proven to be weaker than expected, it led to the bilinguals' average score of 0.55 in both conditions no longer being significantly above chance, which meant comparison with monolingual performance was not possible with this subset of data.

3.4.3 Discussion

There was a developmental effect in that adults were above chance in both critical conditions but children in stressed condition only. This might be due to the fact that the inference is relatively weak and is reinforced by the additional cue of prosodic focus

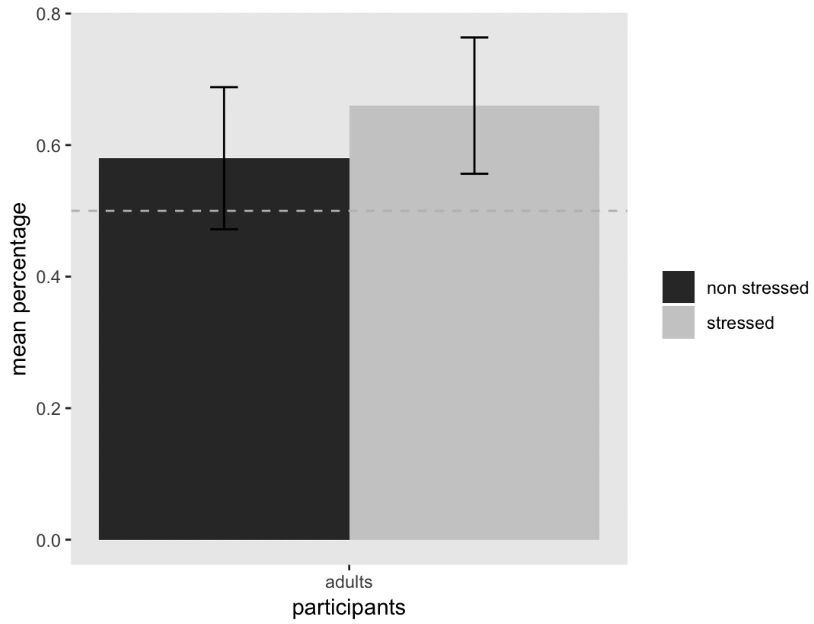


Fig. 3.6 Experiment 1: adults.
non-stressed=modifier without prosodic stress (e.g., 'wet gorp'), stressed=modifier with prosodic stress (e.g., 'WET gorp'). Error bars show 95% confidence intervals

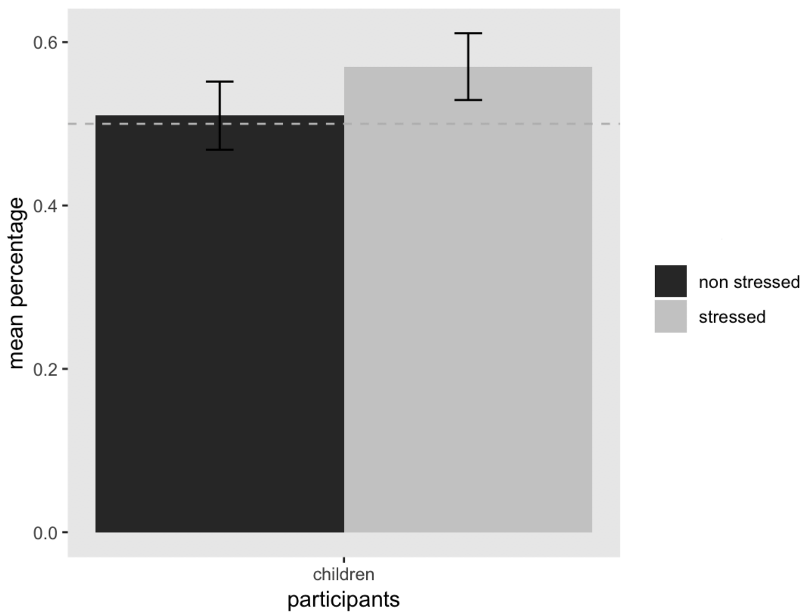


Fig. 3.7 Experiment 1: children.
non-stressed=modifier without prosodic stress (e.g., 'wet gorp'), stressed=modifier with prosodic stress (e.g., 'WET gorp'). Error bars show 95% confidence intervals

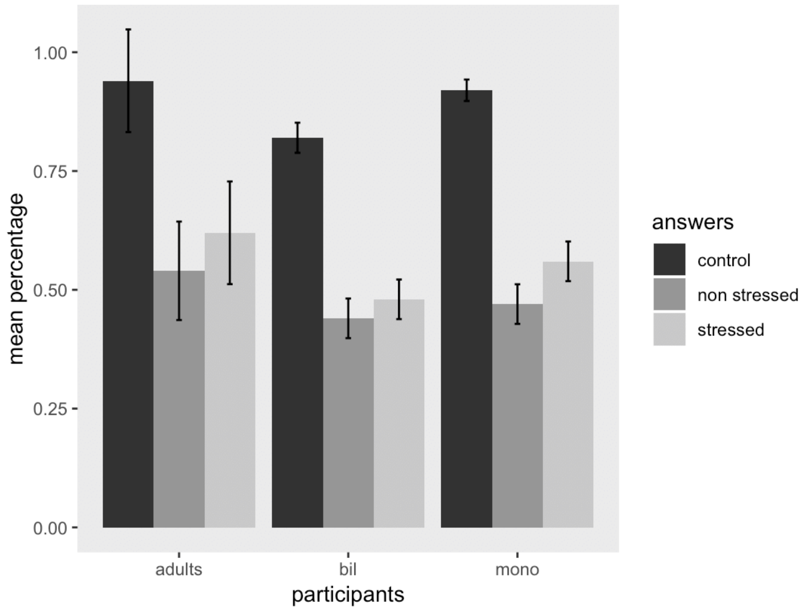


Fig. 3.8 Experiment 1: bilinguals and monolinguals.
control=one semantic match, non-stressed=modifier without prosodic stress (e.g., 'wet gorp'), stressed=modifier with prosodic stress (e.g., 'WET gorp'), bil=bilinguals, mono=monolinguals. Error bars show 95% confidence intervals

which helps direct children's attention towards the use of the modifier, making it more ostensive and raising its informative potential.

One important assumption to keep in mind for the target contrastive inference to be possible is that the speaker is presumed not to be aware that the listener does not know the meaning of the novel words used. Indeed, if this were not the case (if the speaker was presumed to be aware they were using words unknown to the listener), there could be no reasoning about alternatives leading to the inference. In this case, the listener could not reason that 'if the speaker were referring to one of the single aliens they would have used a non-modified noun (e.g., *'gorp'*)' since the speaker would have known this to be unhelpful to a non-knowledgeable listener.

Since vocabulary scores and SES were both shown to significantly impact target choices, it was unsuprising to find that the bilingual group (who were younger, with lower vocabulary scores and lower SES) was performing significantly worse in control

Table 3.1 Experiment 1 and 2: summary of results

Condition	Adults		Monolinguals		Bilinguals	
	mean	SD	mean	SD	mean	SD
(exp1) contrastive (control)	0.92	(0.29)	0.92	(0.27)	0.83	(0.38)
(exp1) contrastive (non stressed)	0.54	(0.49)	0.47	(0.50)	0.44	(0.49)
(exp1) contrastive (stressed)	0.64	(0.47)	0.56	(0.49)	0.47	(0.50)
(exp2) emotional (neutral affect)	0.61	(0.49)	0.65	(0.48)	0.60	(0.49)
(exp2) emotional (negative affect)	0.71	(0.46)	0.84	(0.36)	0.78	(0.41)
(exp2) emotional (positive affect)	0.36	(0.48)	0.60	(0.49)	0.58	(0.49)

Note: raw scores for the contrastive task include all types of errors, i.e., chance = 0.25

than the monolingual group. However, when this delay in structural language (and potentially, given their young age, general attentional skills and ability to focus on the task) was accounted for by examining target answers in both critical conditions relative to perfornance in control, a signicant interaction between bilingual status and condition was found, with bilingual children performing significantly better than expected from control (‘doing more with less’) compared to monolingual children. This means that, compared to the regular use of structural language and semantics for reference resolution in control, bilinguals were particularly sensitive to and efficient at using the pragmatic cue (i.e., the use of a modifier) to reason about the speaker’s communicative intentions in providing that cue. This shows that the differences in exerting pragmatic skills between bilinguals and monolinguals might go further than a simple ‘pragmatic bias’ whereby this particular type of cue is preferred to other types such as semantic content or object similarity. The closer-to-control performance in bilinguals in the non-stressed task could simply be due to their worse performance in control combined with chance performance in critical condition, however this does not apply to the stressed condition given that both bilingual and monolingual groups were shown to be reliably above chance in selecting target in this condition.

3.5 Experiment 2: emotional affect

The purpose of this experiment was to examine the performance of bilingual and monolingual children in a fast mapping task disambiguating the referent of a novel word where learners had to use a similar type of cue as in our first experiment (ie., a prosodic cue) but where there was no need to reason about communicative intentions in providing this cue and, as in previous studies (Yow, 2015; Yow & Markman, 2015), the cue could be used directly as it raised the salience of one referent above the other. In this task, children were shown a display with two novel objects and heard a prompt containing a novel word (eg., *'Oh, look at the dax! Have you seen the dax?'*. They then were shown the two same objects, one of which had been negatively or positively modified (ie., broken or dirtied, decorated with star or coloured pink, etc.) and heard another prompt with the novel word this time uttered with sad, happy or neutral emotional affect (eg.,'(sad voice) *Oh look at the dax! Can you touch the dax?*). This word learning inference relies on the assumption that the speaker should not be overinformative, ie., that they should not use a modifier or lexical stress if they are not required to in order to uniquely identify the referent We predicted that there would this time not be any differences in performance between the monolingual and bilingual groups, since this task required simply to pair an emotional affect with a referent which had been made more salient by said affect, and not to reason about the communicative intentions behind providing the cue, ie., the cue was self-explanatory or self-sufficient and there was no need to reason about its informativeness. Given previous literature on the development of pragmatic competence, we also predict a developmental effect, ie., that adults would perform better than children.

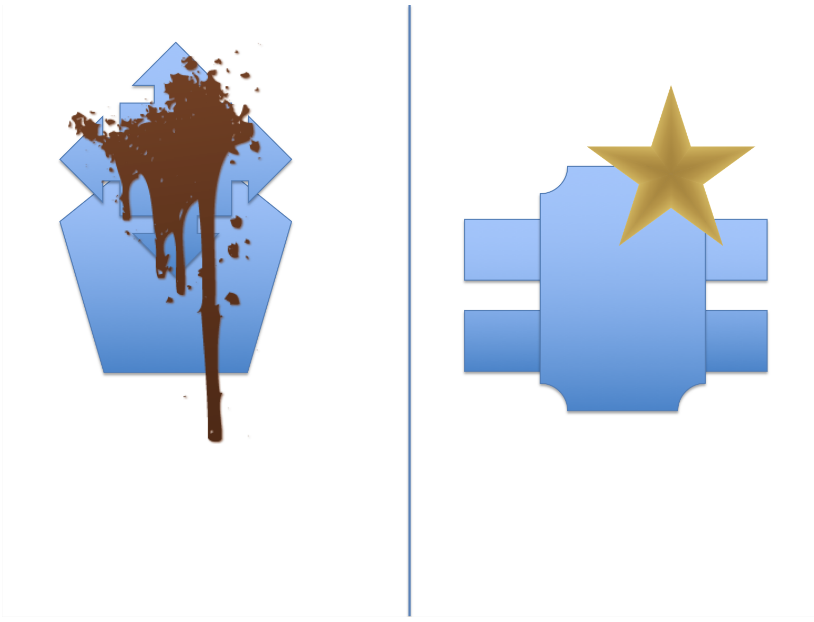


Fig. 3.9 Example stimulus experiment 2, *'(sad/happy/neutral voice) Can you touch the figoo?'*

3.5.1 Method

Participants

Participants were the same as in Experiment 1.

Stimuli

The task was adapted from Berman et al. (2013). Stimuli were 12 pictures, each displaying two novel objects, one damaged object (i.e.,dirtied or broken: mud splash, green splash, hole or dismantled parts) and one enhanced object (i.e., featuring flower, star or brightly lit up). An example stimulus can be found in figure 3.9. A full list of stimuli (utterances and pictures) can be found in Appendix 2.

Procedure

For adult participants stimuli were presented using the Qualtrics platform. For children participants stimuli were implemented and presented on a touchscreen laptop using Superlab 5.0 (Cedrus, San Pedro, CA). Clicks were recorded as raw X and Y pixel coordinates of the point where the screen had been touched and answers were subsequently coded by matching the coordinates to the corresponding answer area among the two possible choices (left or right). Children were tested in a quiet room at their school. The task was part of Mr. Puppet's story and his meeting with aliens who played with strange new objects and damaged some of them but also made some of them 'look prettier'. Children first completed two test trials, one mutual exclusivity trial (to ensure they paid attention to the linguistic information) and one to test their understanding of the task. In each trial they were presented first with the two novel objects in an unaltered state and prompted with *'Oh! Look at these, have you seen these?'* then the altered objects along with an instruction of the type *'Oh! Look at the nurmy, can you touch the nurmy?'* recorded with a positive, neutral or negative voice (emotional affect). The instructions were recorded by a female native speaker of English in a soundproof room using a Sennheiser ME 64 cardioid microphone connected to a Tascam HD-P2 Compact Flash Audio Recorder. Recordings were made in 24bit mono with a sample rate of 44.1kHz and checked for the prosodic and amplitude features characteristic of each type of affect (in addition to the adult study serving as control). Experimental design was within-subject with four trials in each condition (positive, neutral and negative, 12 in total) using twelve novel words (*spoodle, nurmy, goti, fopal, figoo, dazeel, grof, pilk, zarp, mook, slarp* and *klem*). There were two lists of items counterbalanced for word/object pairings and target position. Trials were randomised.

3.5.2 Results

Final results can be seen in Table 1. A logistic regression mixed model with subject and items as random effects for adults showed that negative responses (i.e., dirty or broken object) were significantly above chance in control condition (reference level) (m=0.61, sd=0.49, est=0.52, se=0.15, z=3.35, p=0.0008), significantly higher in negative than control condition (m=0.71, sd=0.46, est=0.47, se=0.18 z=2.55, p=0.01) and significantly lower in positive than control condition (m=0.36, sd=0.48, est=-1.19, se=0.18, z=-6.52, p <0.0001). The regression model for children was again a maximal random effects model adding all design-relevant effects as fixed and subject and item as random effects, random slopes being removed only when leading to non-convergence. Results showed that negative responses were above chance in control condition (m=0.62, sd=0.48, est=0.69, se=0.20, z=3.46, p=0.0005) significantly higher in negative condition than control (m=0.81, sd=0.39, est=1.18, se=0.28, z=4.11, p <0.0001) but not in positive condition (m=0.59, sd=0.49, est= -0.05, se=0.28, z=-0.18, p=ns). There was also a significant effect of vocabulary scores with higher scores leading to a negative bias in control (est=0.20528, se=0.08591, z=2.389, p=0.0169). No other effects were significant, the bilingual group performance did not differ from the monolingual group. A graph showing the results for the adults can be seen in figure 3.10, and for children in figure 3.11. The results broken down for bilingual and monolingual groups can be seen in figure 3.12.

3.5.3 Discussion

There was a negative bias in the neutral affect control condition found in both children and adults whereby negative choices (i.e., dirty or broken objects) were preferred to positive ones (i.e., enhanced/decorated objects). This default preference for a negative interpretation or outcome is found in the literature and potentially results from a higher

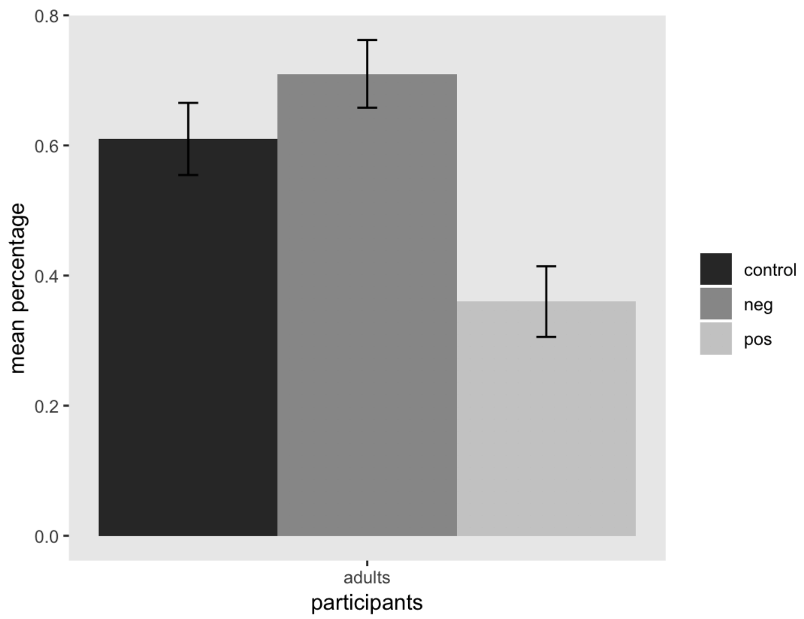


Fig. 3.10 Experiment 2: adults.
control=neutral affect, neg=negative affect, pos=positive affect. Error bars show 95% confidence intervals

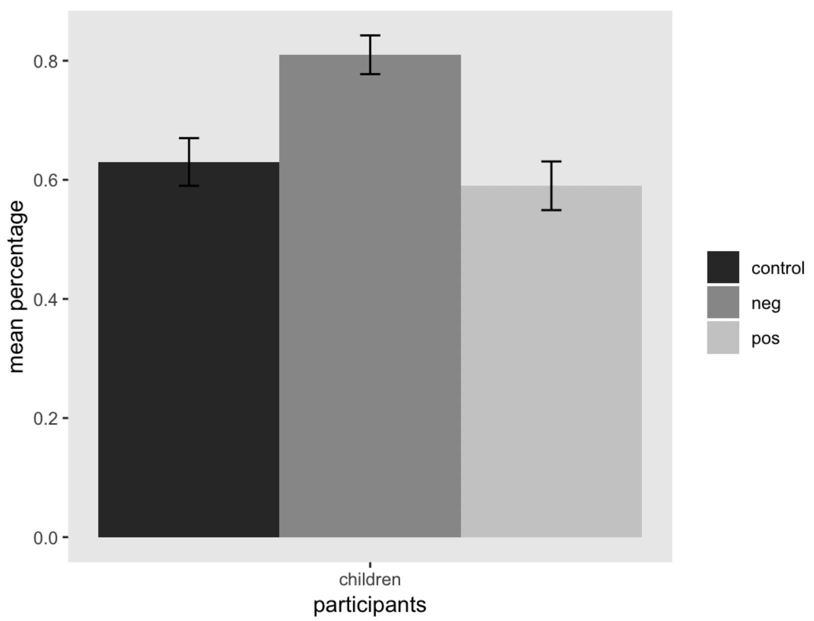


Fig. 3.11 Experiment 2: children.
control=neutral affect, neg=negative affect, pos=positive affect, bil=bilinguals, mono=monolinguals. Error bars show 95% confidence intervals

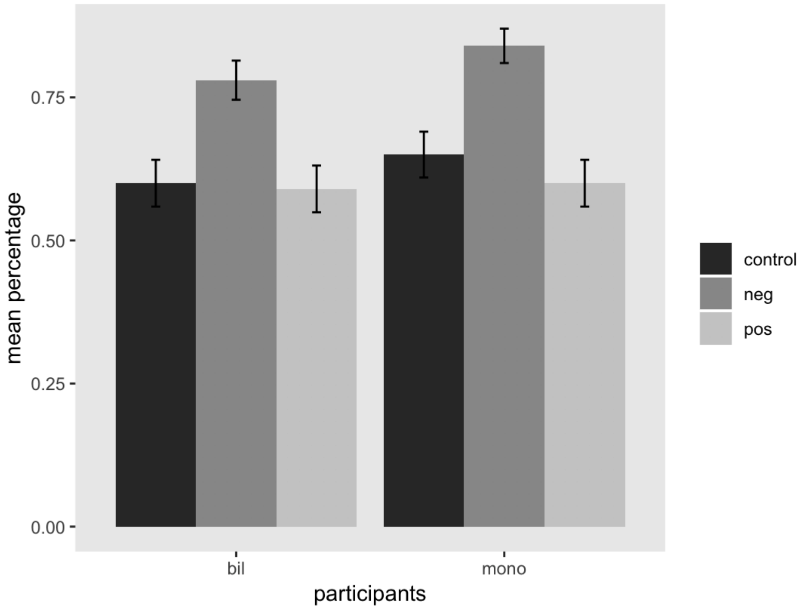


Fig. 3.12 Experiment 2: bilinguals and monolinguals.
control=neutral affect, neg=negative affect, pos=positive affect, bil=bilinguals, mono=monolinguals. Error bars show 95% confidence intervals

salience of negative versus positive events (Berman et al., 2013), which is why answers in the negative and positive conditions were compared to performance in the control condition rather than to chance performance in the regression model. There was also a developmental effect, with adults performing as expected in both critical conditions (i.e., making significantly more negative choices when hearing a new label uttered with negative than with neutral and positive affect, and significantly less negative choices when hearing a new label uttered with positive than with neutral and negative affect), whereas children were succeeding only in the negative condition and seemed to fail to perceive or use positive affect to direct their referent choices in the positive condition. This is again something that has previously been found and is potentially, like the negative bias mentioned above, the result of a higher salience of negative versus positive. Just as negative events appear to be more salient that positive ones, negative prosody or emotional affect might have higher salience than positive, which would lead to children being able to interpret and use it earlier (Berman, Chambers, &

Graham, 2010; Nelson & Russell, 2011). The higher salience of negative events and affect appears rational from a biological/survival instinct point of view, since ignoring warning about bad outcomes might have direr consequences than ignoring good news.

No interaction between bilingual status and performance in critical condition relative to control condition was found this time. However, performance in this task relied entirely in being able to recognise and associate the specific valence of the cue (negative or positive) with the corresponding event (negative, i.e., damaged object or positive, i.e., enhanced object). Since the prompts always had the same linguistic form, there was no advantage of having a stronger vocabulary knowledge of modifiers as in the first task, and contrary to specific language exposure, there is no principled reason why bilingually exposed children should have had more or less experience with or in interpreting different types of emotional affect. More importantly, unlike in the contrastive task, success in this task did not require reasoning about *why* the speaker was providing a certain cue since emotional affect, contrary to prosodic stress, has an intrinsic value or valence (negative or positive) which can be directly linked to the corresponding (damaged or enhanced) referent. We further develop this explanation in the general discussion.

3.6 General discussion

The goal of experiment 1 and 2 was to investigate monolingual and bilingually exposed children's use of pragmatic cues and prosodic cues for reference resolution of a novel word. Past literature had examined bilingual and monolingual children's performance in reference resolution tasks which involved using a socio-pragmatic cue such as pointing, eye gaze or emotional affect and found bilingual children generally performing better than monolinguals (Yow et al., 2017; Yow & Markman, 2015). However, given that the experimental paradigms used required ignoring an irrelevant cue (e.g., semantic meaning, object similarity or object salience) in favour of the pragmatic one, it remained unclear whether these results were driven by attentional biases (i.e., more attention towards these types of cues or a preference over other types of cues) or by the activation of pragmatic reasoning per se (i.e., reasoning about speaker's communicative intentions). To examine the latter hypothesis, we tested monolingual and bilingually exposed children aged 4 to 6 years old in two tasks using different types of prosodic cues for reference resolution of a novel word: contrastive stress (e.g., 'Touch the WET gorp') and emotional affect (e.g.,'*[sad tone]* Oh look at the figoo, can you touch the figoo?') where success did not rely on ignoring an irrelevant cue and could thus not be solely the result of better inhibitory control abilities.

We replicate previous findings that both adults and children are able to use negative affect to choose between damaged and enhanced referent for a novel label but young children around the age of five do not seem able to reliably make use of positive emotional affect for the same purposes (Berman et al., 2013). We demonstrate that prosodic stress does significantly modulate performance of a contrastive inference for novel word fast-mapping in children as well as in adults (an open question, cf. Kronmüller et al., 2014), with children appearing to be unable to reliably perform such an inference when the use of a contrastive modifier was not emphasised by contrastive stress. This is in line with previous work with adults demontrating a facilitating effect of intonational focus for contrastive inference in referent resolution (Sedivy et al., 1999). We also find a gender difference, with males producing less pragmatic answers than females, an effect which has been documented before (Stiller et al., 2015).

We found bilingually exposed children to perform significantly better than expected in the contrastive stress task (i.e., close to monolingual performance despite being younger, with lower SES and vocabulary levels) but not in the emotional affect task (where the ability to overcome a negative bias was as expected from demographic variables in both groups). This indicates that bilingual exposure enhances not only the attention to socio-pragmatic cues (as demonstrated in previous studies) but also the ability to reason about speaker's communicative intentions in providing the cue. Indeed, the main difference between the two types of prosodic cues in our tasks was the following: while contrastive stress required pragmatic reasoning to be interpreted (since it had no intrinsic valence or meaning, i.e., the hearer will know that the modifier 'wet' has been emphasised but since there are two wet aliens this by itself will not be enough to resolve reference without asking *why* it was emphasised), emotional affect on the other hand intrinsically contained enough information for reference resolution by simple association (i.e., pairing negative emotion with negative event). As we have said, this can be related to Grice (1975)'s dichotomy between natural meanings and non-natural meanings, the difference being the need to recognise a communicative intention. The absence of a number of significant effects found to impact performance in the first task (such as gender or SES) further emphasises that the two tasks involve different sets of skills. While the finding of a bilingual performance closer to control than monolinguals' could be due to the former performing worse in control and at chance in the non stressed condition, we do not believe this is the case in the stressed condition, given that both bilingual and monolingual groups performed significantly above chance in this condition.

In conclusion, we presented the results of two tasks investigating monolingual and bilingually exposed children's use of pragmatic inference for fast mapping a referent and novel word. We found the bilingually exposed group to perform above expected levels compared to monolinguals in the first task, where above chance performance could only be achieved by using a prosodic cue to reason about speaker's communicative intentions. We conclude that this work provides evidence for differences in the processing of pragmatic cues by bilingual and monolingual children which are not due to better inhibitory skills or to a generally higher sensitivity to social cues, but to performing true pragmatic inference by reasoning about communicative intentions in the context of word learning. We propose that these differences are the result of adapting to the challenging aspects of a bilingual learning environment, such as higher risks of miscommunication and the need to efficiently and quickly acquire words from complex and variable input. This is therefore another instance where the highly adaptive nature of child cognition is evidenced, a case of 'doing more with less'. Moreover, empirically validating the distinction between using social cues and reasoning about intentions underlying social cues is a fundamental cornerstone of pragmatic theory and might help provide insights about separate developmental timelines for pragmatic competence.

Chapter 4

Frequency inference and defaults

I'm like that. Either I forget right away or I never forget.

— Samuel Beckett

4.1 Pragmatics and costs

Consider again the classic scalar implicature example :

(1) Some [*but not all*] of the students passed the test.

And compare it to the following exchange (Grice, 1975:32):

(2) A: Smith doesn't seem to have a girlfriend these days.
B: He has been paying a lot of visits to New York lately.
[*Smith probably has a girlfriend in NY*]

B's answer in (2) constitutes a typical example of implicature in a Gricean sense: it arises from a specific context, probably through a process of conscious inference, it involves reasoning about what the speaker wants to communicate, it is detachable (it would arise if using slightly different phrasing), cancelable, and global (it arises from processing the entire utterance).

The implicature in (1), on the other hand, appears to be of a different nature. While the implicated meaning in B's answer only arises in a specific context (taken in isolation, without the previous utterance, it would not carry the implicature), the one in (1) appears to arise *in the absence* of a specific context which would cancel it, to be more closely related to a particular word (to arise *locally* rather than *globally*) and to require less, or perhaps no conscious inference. This prompted Grice to create a category of GCI (*Generalised Conversational Implicatures*) which would be generated *by default*, that is to say, normally, in the absence of special circumstances, by 'the use of a certain form of words in the utterance'. The idea of a category of default local implicatures was endorsed by neo-Griceans such as Levinson (2000) and Horn (2004), while relevance theorists (Sperber & Wilson, 1986; Carston, 1995) have insisted on implicatures being a property of whole sentences or propositions and nonce ('one-off') generated. Similarly, as we mentioned in the previous chapter, theories differ as to whether the default meaning is generated first (Levinson, 2000) or only after the literal meaning has been processed (Huang & Snedeker, 2009).

This debate has a number of implications for theories of meaning. The issue with GCIs comes when trying to separate *what is said* from *what is implicated*. Such a distinction is useful both from a speaker's point of view ('Can X be held accountable/said to have lied about I?') and from a hearer's point of view ('Which part of the meaning can be considered as being directly conveyed through the words themselves?'). However, it seems impossible to find a criterion to operate this separation in a definite way. Truth conditional status seems to cut across categories: conventional implicatures (e.g., '*He is English but brave*') do not participate to propositional content yet are explicit, whereas some GCIs have truth conditional effects ('*If the old king died of a heart attack* *and a republic was declared, Tom would be happy, but if a republic was declared and the old king died of a heart attack, he would not'*, Cohen, 1971). Defaultness (being derived automatically) and locality (being tied to a specific lexical item) would tend to suggest that such meanings are part of the explicit content, however neo-Griceans such as Horn (2004) would argue that this category cannot include cancelable meanings.

A number of authors substituted the classic dichotomy with a tripartite distinction: sentence, utterance-type, utterance-token (Levinson, 2000); explicature in Relevance Theory (Carston, 2004), impliciture in Bach (1994). These categories seem to be based on two main criteria: 1) is the meaning *additional* to sentence meaning and 2) is it available *consciously*? They are particularly explicit in Recanati (2004) distinctions between primary and secondary meanings, and between 'saturation' and 'free enrichment'. Consider again Bach (1994)'s example, uttered by a mother whose child has just injured their knee:

(3) You [*the child*] are not going to die [*from this cut*].
[*Stop making a fuss*]

Recanati defines primary pragmatic processes as non-inferential, associative-like mechanisms providing the explicit content of the utterance, whereas secondary pragmatic processes produce the implicit meanings/classic conversational implicatures. Thus, in (3), the enrichment [*from this cut*] is primary meaning whereas the final [*Stop making a fuss*] interpretation is secondary meaning. Recanati also distinguishes in the primary processes between 'saturation' (for non-truth evaluable propositions, e.g., the enrichment of the pronoun *you*) and 'free enrichment' (e.g., the enrichment of the verb *die*).

However, there does not seem to be a principled distinction to be made between cases of 'saturation' such as (4) and cases of 'free enrichment' such as (5), where the difference in truth evaluability status seems at best incidental, at worst irrelevant, and in any case delicate to establish (could the broadly understood (4) not be considered truth evaluable?)

(4) The fence isn't strong enough [*for anything*]

(5) I haven't eaten [*ever*].

Similarly, the distinction between primary and secondary meanings does not appear, in fact, to align with an inferential criterion.

(6) John is a doorbell.

(7) A: Is John married?

B: He has a wife. [*Yes*]

In (6), complex inference will be required to arrive at primary meaning, whereas the full secondary meaning of B's answer will require little or no inference. Thus, what primary and secondary meanings seem to recover are the differences between steps involving *what* has been said and *why* it has been said (i.e., which speech act or goal is achieved through the utterance, e.g., getting the child to stop complaining in (3) or answering A's question in (7)).

Are there meanings derived by default? The difficulty resides in how these are defined. It seems difficult to define them as context-independent (cf. Levinson, 2000), in the view of examples such as (11), where the additional meaning does not arise without there being a contradictory context (Carston, 1995). Similarly, while the meanings of 'fish', 'financial institution' and 'some not all' might arguably be more salient or frequent for a particular item in (8), (9) and (10), they do not seem to qualify as the default meanings in these utterances.

(8) (*pointing at man*) John is a shark.

(9) The water flowed down the bank and into the stream.

(10) If some of the students passed, the teacher will be happy.

(11) I did the dishes and gave the baby a bath.

One solution is to define defaults as meanings that are automatically derived in a given situation of discourse (Jaszczolt, 2009), whereas salient meanings are simply the most frequent, familiar or stereotypical ones (Giora, 2003). According to Jaszczolt's *Default Semantics*, utterance meanings are obtained by combining the output from a variety of sources: word meaning and sentence structure (WS), situation of discourse (SD), properties of the human inferential system (IS), stereotypes and presumptions about society and culture (SC) and world knowledge (WK). However, in this theory an utterance such as (13) would be processed defaultly (just as (9)), whereas (14) would involve a 'conscious pragmatic inference' (CPI) to arrive at an attributive rather than referential reading. Yet both (13) and (14) seem to involve a similar case of something akin to an instance of 'garden-pathing'. It is also dubious whether the sources of knowledge need to be distinguished in so many categories, for example (e.g., cultural versus world knowledge), or could be subsumed under 'hearer's knowledge at given time'. Perhaps a more crucial distinction could be the one between common and privileged ground (Keysar et al., 2000), which we come back to in the next paragraphs.

What are defaults and why would some meanings be defaultly processed? As we said earlier, humans are lazy. Some processes have evolved to be more or less default and automatic, either as innate biases or from habit and experience, whereas others, more controlled and costly, require reasoning (Kahneman & Egan, 2011). Defaults exist to reduce costs. So when *is* meaning interpretation costly? In language processing as in other domains, there is a strong tendency to go for the path of least resistance. In the search for speaker meaning (or assessment of potential mappings between linguistic expressions and referents), hearers will access both sources of information and referents as they are ranked, i.e, on the basis of their availability. Consider utterances (12)-(16).

Why are these utterances likely to be more costly/not proceed defaultly, compared to (8)-(11)?

(12) John is a doorbell.

(13) The bank on which the water flowed was very steep.

(14) The first daughter to be born to Mr. and Mrs. Brown will be called Scarlett.

(15) He opened the door and she handed him the key.

(16) That professor is great.

In all above cases, the cost seems to arise from the intended meaning not being sufficiently available *at the time of processing the relevant items.* As *Default Semantics* and constraint-based theories suggest, we derive meaning through the integration of multiple sources of information. Because of the temporal nature of language and constraints on attentional window (e.g., a limited amount of visual information is being processed at one time), we do this in a continuous and incremental way, taking into account *number*, *strength* and *convergence* of cues. At any given point in processing *there is* a default interpretation, which is the most available both in terms of *source* and *referent.* For the intended meaning to be processed with little cost, it needs to have been made more available or salient than other meanings by the cues processed so far. Otherwise, costs will arise, both because some type of inference will be needed, and because the default reading will have to be inhibited.

In (8), while the intended meaning of 'shark' is not technically literal, it is available enough, as a conventional metaphor, once 'John' has been processed (and thus made a human-compatible meaning more probable). Similarly, in (10), the presence of the conditional clause introduced by 'if' is a strong enough cue for a non-upper bounded meaning of 'some', which is available as a relatively frequent alternative to the pragmatically enriched 'some'. On the other hand, in (12), the presence of 'John' is not enough to make available a human-compatible meaning for 'doorbell' to overcome its literal meaning, as such a meaning of 'doorbell' might never have been encountered before and will require an inference. This inference may then be performed in an associative manner (i.e., 'What are doorbell traits that can be attributed to a human?') or by performing intention reading (i.e.,'Why would the speaker compare John to a doorbell?'). In (13), as opposed to (9), the intended meaning of 'bank' has not been previously made more salient by 'the water flowed down' and thus is not available enough compared to the more salient meaning 'financial institution'. The same problem arises in (14) and (15), though it seems harder to pinpoint the characteristics of the previous context which make available an enriched meaning of 'and' in (15) (as it turns out probably wrongly) and not in (11) (perhaps an event of opening a door leads to expect something which temporally follows, whilst an event of doing the dishes does not?). In (16), the cost comes from the fact that an ironic reading is less often readily available (apart from very conventionalised forms) and will require speaker-specific inference. Importantly, costs are not linked to type of process which led to an expression having a specific meaning, but to whether an inference is required *online*. For example, scalar implicatures such as 'some but not all' might progressively become more default, which means they will be easier to process *in certain contexts* and might require less inferential effort, but they can still be reconstructed as having been calculated from assumptions of informativeness (just like the process of creating the 'shark' metaphor could be reconstructed using the same assumptions, Kao, Bergen, & Goodman, 2014).

In this view, the question of whether literal meanings come first, or whether pragmatic inferences are costly, becomes irrelevant. As Hanks (2004) says: *'There are no literal meanings, only varying degrees of probability'* (or, rather, varying degrees of availability). Instead, Hanks distinguishes between 'norm' (or retrieval) and 'exploitation of norm' (or inferential processes). Idiom processing constitutes a good example of this balance. Indeed, for a long time views diverged on how idiomatic expressions were processed, between theories which considered them to be ‘long words’ simply retrieved as such in memory (Chomsky, 1980; Gibbs Jr & O’Brien, 1990) and compositional approaches driven by the evidence of syntactic flexibility and individual meaning effects, which did not systematically end when the ‘idiomatic key’ (or point of recognition of the idiom) had been reached. According to the compositional perspective, idioms were ‘nonce’ built from individual meanings and immediate context (Gibbs Jr, Nayak, & Cutting, 1989; Nunberg, 1987) . However, idioms benefit from certain properties which do seem to distinguish them from other sequences (e.g., idioms which are ‘decomposable’, that is to say whose logical form can be mapped to their meaning have an advantage in processing speed, both over non-decomposable idioms e.g., Titone & Connine, 1994, as well as over non-idiomatic speech and novel metaphors, Caillies & Declercq, 2011), which differ in their ‘literal’ status but not in their inferential versus retrieved status). In light of this, it appears once again that only models which integrate both the analytic/inferential and retrieval/memory outputs together, such as the *hybrid model* (Titone & Connine, 1999) and *construction-integration model* (Caillies & Butcher, 2007; Kintsch, 1988) can accurately predict such effects.

Thus, rather than different strategies, or different sources of information/constraints overriding each other, we can see meaning derivation as a probabilistic combination of default and inferential processes, of retrieval and reasoning.

4.2 Cues, constraints and development

As we have seen, some interpretations are likely to be more costly than others. This will depend on cue *number*, *convergence* and *strength*. That is to say, the easiest, default interpretations will be the ones that have many strong cues and are unchallenged by salient competitors. Cue strength is determined by a number of features which define availability: salience (e.g., visual, memory based, etc.), type of access (retrieval or inferential) and experience. The strongest cues are then those that give the most direct and easy access to meaning, with visual access likely to be be more direct than retrieval in memory, direct retrieval being easier than associative, associative easier than inferential, and inferential based on egocentric knowledge easier than that requiring perspective-taking. Thus, deriving the meaning of a word used with a salient visual referent present (e.g., *'the cat'*) will be easier than retrieving a representation from memory. Similarly, associating two meanings (e.g., *wife* mentioned in answer to a question about marriage) might be easier than inferring from a single known meaning (e.g., enriching 'some' based on calculations about informativeness and alternative expressions available). Finally, inferring meanings based on speaker-general assumptions of informativeness should be easier than inferring meanings based on perspective-taking, as in cases of irony such as *'That professor is great'*, since speaker-general knowledge can be modelled on egocentric knowledge, and is thus more available than speaker-specific knowledge.

Where do pragmatics lie? If defined as contextual contributions to linguistic meaning, their status is uncertain, as they will most likely cut across the distinction between default and costly processes, from automatic enrichments of individual words (pronouns, etc.) to complex inferences based on the relevance of a particular utterance in relation to the broader discourse context and speaker goals. But if we define pragmatics as reasoning about informativeness, then pragmatic inferences will tend to fall on the costly side of the default/resource-demanding spectrum. As such, they will be used in an active manner when more default heuristics which filter obviously attractive or implausible meanings have failed to provide the hearer with a straightforward answer. Among inferences based on speaker-general knowledge, as we said earlier, the ones which are based on reasoning about alternative expressions (as opposed to just alternative referents, as in size or frequency based inferences), such as ad hoc and contrastive inferences are probably stronger and less costly since they benefit from mutual reinforcement/multiple cues: in the 'hat and glasses' example from Stiller et al. (2015), the target expression e1 (e.g., 'glasses') is more specific for the target referent r1 ('glasses') AND in addition the alternative expression e2 ('hat') is specific to the alternative referent r2 ('hat and glasses'). In contrast, in the frequency inferences we describe in the next section, there is no reasoning about alternative expressions (of the type 'if they had wanted to refer to alternative referent, they would have said X').

Thanks to defaults, humans get to be lazy in communication: associations and inferences do not have to always be computed online, and thus speaker model does not have to be continuously actively updated either. This behaviour is rational and efficient, since default heuristics (such as assuming a fixed speaker model, shared common ground or relying on most cognitively available descriptions) apply successfully in a vast majority of cases. However, this laziness comes with a price: it can sometimes lead to communication failures and means that pragmatic reasoning is not quite optimal, in comprehension or in production. Because people firstly rely on default mechanisms and are not continuously reasoning actively about informativeness or updating their model of the speaker, they for example fail to immediately notice non-optimally informative descriptions (Engelhardt, Bailey, & Ferreira, 2006; Rubio-Fernández, 2016) or to adapt their own in production (Horton & Keysar, 1996; Lane & Ferreira, 2008).

An example of the importance of defaults in experimental paradigms might lie in the puzzling diversity of results in studies of scalar implicature using different paradigms. Syrett, Lingwall, et al. (2017) and Syrett, Austin, et al. (2017), which we mention in Chapter 2 tested children aged 3 to 6 years old on scalar implicatures. In these (as in other studies), children (bilingual or monolingual) did not appear to perform the scalar inference, and the form of the instruction had a major impact on children's rates of pragmatic answers, with instructions mentioning the items of the scene biasing towards a positive answer not only in pragmatic infelicitous cases (*'Some horses have carrots'* when all horses have carrots) but also in semantically incorrect cases (*'All horses have carrots'* when only some do), particularly for bilingual children (37% positive answers to semantically incorrect sentences). This is similar to findings in (Horowitz et al., 2018; J. Sullivan, Davidson, Wade, & Barner, 2017) When a neutral instruction was used (*'Did she do well?'* in Syrett et al., 2017a) the effect disappeared. This is probably due to lack of experience with quantifiers, which means their meaning is not strong enough to inhibit a tendency to answer an easier question (e.g., *'Are there horses and carrots?'*) and produce a positive answer based on the nouns only (cf. Horowitz et al., 2017). This effect might be particularly strong in bilinguals since their vocabulary knowledge and experience using quantifiers of each language is even weaker than monolinguals' and could lead to an even stronger tendency to ignore quantifiers altogether in favour of an 'easy' answer using nouns only. These various methodological caveats make it difficult to interpret the result of these studies in terms of tasks that would purely assess pragmatic competence in bilinguals, rather than world knowledge, metalinguistic awareness or noun bias.

Does this mean people choose one strategy or type of cue over the others as default and do not consider other, less readily available sources of information? Evidence would in fact seem to indicate that all 'channels' of information stay open, but the amount of attention given or effort spent on them varies. An example is precisely the debate over the integration of common ground knowledge. It has been an open question whether hearers consider common ground knowledge from the very beginning of utterance processing, or rely first on egocentric knowledge *by default*, engaging in perspective-taking or speaker modelling (i.e., Theory of Mind) only when failure occurs. Indeed, it would seem that the interpretation of a definite referring expression such as *'the cat'* should be restricted to the shared ground between two speakers. Imagine a friend telling a story and uttering *'the cat'*. It is unlikely that you will interpret this as referring to your neighbor's cat who got stuck atop a tree once and whom your friend has never heard of. Yet, in an influential study, Keysar et al. (2000) showed that listeners did not restrict their attention only to objects in the common ground.

In this study, Keysar et al. used a paradigm where a speaker (the director) and a listener sat across each other in front of a wooden array formed of cubicles containing objects. The director would give instructions to the listener to move objects (e.g., *'Move the big candle above the funnel'*). Some of the cubicles were closed, so that the objects inside them were visible to the listener but not to the speaker. In critical trials, there were two objects which satisfied the target referring expression (e.g., one very big candle, one big candle and one small candle) but the best match for the target expression (i.e., the very big candle) was hidden from the speaker and thus should not have been taking into consideration, yet listeners first directed their eye gazes at the distractor object, and sometimes even reached out to it. Keysar et al. thus concluded that listeners started with interpreting referring expressions according to egocentric perspective.

However, other studies obtained contradictory results, that is, listeners integrating common ground from the earliest moments of processing. In Hanna, Tanenhaus, and Trueswell (2003) and Nadig and Sedivy (2002), when the privileged and common ground matches were equally perceptually salient (e.g., two red triangles), listeners did show early preferential looking towards, and faster processing of, objects in the common ground.

One way of accounting for these results is to take a constraint-based approach, similar to that of *Default Semantics*, where interpretation results from the combination of outputs from a variety of sources. This is the approach taken by Hanna et al., (2003), who argue in favour of a *common ground constraint*, which favours objects in the shared ground and *perceptual constraint*, which favours perceptual matches. These constraints would then explain why listeners readily processed shared matches in Hanna et al. (2003), while they did not in Keysar et al., as the output of the two constraints (shared match and best perceptual match) diverged in the latter study. However, as pointed out by Heller, Parisien, and Stevenson (2016), the results from a third study show that this cue integration has to proceed in a *simultaneous* fashion. Indeed, in Heller, Grodner, and Tanenhaus (2008), listeners were faced with a setup similar to that of Keysar et al., where two contrastive pairs of objects were displayed (e.g., a big and small funnel and a big and small candle) only one of which was in common ground (e.g., the big and small candle were both visible to the speaker, but the small funnel was not). Eye-tracking analyses showed early preferential looking towards the match in the shared contrastive pair, a result which would not be predicted by the output *common ground constraint* or *perceptual constraint* processed separately and *then* weighted and combined (as they would both predict the two big objects to the same extent, being both in common ground and both equal perceptual matches). While other authors have assumed different strategies of perspective-taking to explain these contradictory results, Heller et al., (2016) propose a probabilistic model of simultaneous constraint integration which can jointly account for the apparent behavioural differences.

Within the Rational Speech Act framework, Bohn, Tessler, and Frank (2019) propose a similar Bayesian model to explain how listeners flexibly integrate speaker-general and speaker-specific (common ground) knowledge, the first one as an inference which can be calculated based on assumptions of informativeness and desire to avoid ambiguity, the second as a modified prior to account for referent preferences.

How do children fare in performing this cue integration? Children have reliably been found to struggle in deriving scalar implicatures of the 'some-but-not-all' type (Huang

& Snedeker, 2009; Noveck, 2001) , with studies using different paradigms obtaining varying levels of performance (Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004; Skordos & Papafragou, 2016; Teresa Guasti et al., 2005). These difficulties have been hypothesised successively as the result of a pragmatic deficit, an alternative accessibility issue or, more recently, lack in semantic knowledge of quantifiers (Horowitz et al., 2018).

In the case of pragmatic inferences such as scalar implicatures, one explanation for the difficulties experienced by children resides in the salience or accessibility of alternatives (the *Alternatives Hypothesis*, Barner & Bachrach, 2010; Barner, Brooks, & Bale, 2011). Children, as adults, seem to perform interpretations on the basis of multiple cues related to various constraints and to rank these cues according to their reliability and availability or ease of perception (cf., constraint-based accounts of pragmatic inference, E. Bates, MacWhinney, & MacWhinney, 1987; Degen & Tanenhaus, 2015; Nappa, January, Gleitman, & Trueswell, 2004). However, the strength and weighting of each type of cue in children might differ from adults given that they have less experience with social reasoning and benefit less from processing capacities which would allow them to inhibit salient but irrelevant cues. Given that they benefit both from less processing capacity and less experience with meaning interpretation, perspective-taking and pragmatic processes in general than adults (thus leading to weaker cue strength for these sources of information), it would seem likely that they might experience some difficulties.

4.3 Frequency inference

Following Tenenbaum and Griffith's (2001) *Size Principle*, learners can base their choices of referent for a novel word purely on specificity, i.e., relative frequency or novelty in context, even in the absence of salient alternative expressions more specific to alternative referents (e.g., *'all'* being more specific to the meaning 'all' than *'some'* would be). For example, if presented with a character with one feature (e.g., a dinosaur with bandanna) and another with two features (e.g., a dinosaur with headband and bandanna) described as a *'dinosaur with a dax'* hearers should infer that the novel word refers to the target character's most specific feature, or least frequent feature in context, i.e, the headband rather than the bandanna (Frank & Goodman, 2014). Similarly, children tend to map a new word to the most novel referent in context (Akhtar, Carpenter & Tomasello, 1996). More generally, more frequent referents (such as ceiling, floor, table, etc.) are given a lower probability as potential meanings for a novel word due to their cross-situational ubiquity.

One perhaps important difference between scalar or frequency inference, and the contrastive inferences examined in Chapter 3 is that the first relies on assuming that a speaker should not be *underinformative* (e.g., using *'some'* to mean 'all') while the second relies on assuming that the speaker should not be *overinformative* (e.g., using *'the tall glass'* when they could have said *'the glass'*). There are both theoretical and experimental reasons for thinking that overinformativeness is not seen as infelicitous, or unusual, as underinformativeness. Indeed, the former, unlike the latter, does not lead to referent identification failures and is often witnessed in naturalistic speech, especially for salient properties such as colour adjectives (Rubio-Fernández, 2016; Sedivy, 2003). Moreover, both children and adults have been shown to penalise overinformativeness less than underinformativeness in experimental setups (Davies & Katsos, 2009; Engelhardt et al., 2006). However, in a study by Davies and Katsos (2013) where 3 and 4 year olds were presented with referring expressions which were overinformative both with regard to the visual context and in terms of typicality of modification (e.g., *'the cat with the tail'* when only one normal looking cat is present in the picture), 4 year olds did show signs of confusion, while 3 year olds did not, proving that even young children are sensitive to overinformativeness.

The bilingual effect on deriving the contrastive inference in Experiment 1 was a relative effect which did not reach numerical significance. However, there are reasons to believe that this advantage would be most apparent in a stripped-down paradigm when only one pragmatic cue is present. Indeed, in Rowe et al. (2015), there were no differences between monolingual and bilingual children's extensions of a novel word to a referent which already had a familiar name when two pragmatic cues were given (e.g., *'This is a skiff, it's a kind of boat'* but bilinguals, unlike monolinguals, were above chance in a condition where only one pragmatic cue was given (e.g., *'This is a skiff, it's a boat'*). Thus we hypothesise that the differences between the monolingual and bilingual group might be most apparent when a minimal amount of information is given. However, the inference still needs to be accessible enough for young children to be able to derive it, since, as we mentioned previously, children of this age tend to struggle when pragmatic inference paradigms do not involve additional discourse support for the target interpretation.

While children in Experiment 1, unlike adults, had failed at deriving a pragmatic inference from a single cue (adjectival modifier), this inference was based on assuming that a speaker should not be *overinformative.* As there are reasons to believe that an inference based on penalising a speaker for being *underinformative* might be stronger, this is what we investigate in the following experiments.

4.4 Experiment 3a: frequency inference

Experiment 3a intended to replicate adults and children's performance of a frequency inference for word learning by adapting a paradigm similar to Frank and Goodman's

(2014) as a computerised task. Given the finding that children in experiment 1, unlike adults, did not seem to be able to use a single prosodic cue for word learning when relying on the assumption that the speaker would avoid being *overinformative*, the purpose of this experiment was to investigate whether a single cue was sufficient in the context of assuming that a speaker should avoid being *underinformative* which, as we have seen, is a more strictly penalised violation and thus probably more reliable inference. In this task, children are shown a display with two similar characters (eg., two cats), one with one novel object (e.g., object A), the other with two novel objects, including the same object (e.g., object A and object B). At training time, children hear a prompt using a novel word (eg., *'Oh, a cat with a dax!'*), then at test time they have to choose between two new characters (eg., two frogs), each of which only has one of the objects (eg., a frog has object A and a frog has object B) after hearing another prompt (eg., *'Can you touch the frog with the dax?*). This inference relies on the assumption that describing a specific character using the novel object would be underinformative if both characters have the object (ie., if the speaker is referring to the shared more frequent object A). We predict that children should be better able to perform the inference in this context, and furthermore we predict that, as in experiment 1, under the hypothesis that they benefit from enhanced abilities to reason about informativeness, bilingual children should outperform monolingual children. We also predict a developmental effect, ie., that adults would perform better than children.

4.4.1 Method

Participants

Participants in the adult study were students of the University of Cambridge (n=32) who were recruited through email and completed the study online.

Children participants were the same 270 children as in experiments 1 and 2 (cf. Chapter 3).

Stimuli

Experimental stimuli for the critical trials were 4 pictures featuring 4 characters each with novel objects: 2 training characters (one target character circled in red with target and distractor object, one distractor character with distractor object only) which varied between items (cats, birds, dogs and fish) and 2 test characters, which were always frogs (one with target object only, one with distractor object only). The control condition was a mutual exclusivity condition featuring 4 disambiguation trials with pictures identical to critical items except the distractor character had a familiar object (e.g., an apple) and the target character had an unknown object. The target character was not circled in red in the mutual exclusivity condition. There were two lists counterbalancing which novel object was the target and the distractor. The position of the target test frog relative to the target training character (same/different) was also counterbalanced between items. An example stimulus for the frequency inference condition can be found in figure 4.1, and for the control/mutual exclusivity condition in figure 4.2. A full list of stimuli (utterances and pictures) can be found in Appendix 3.

Procedure

The task was implemented on a touchscreen laptop using Superlab 5 (Cedrus, San Padro, CA) with instructions recorded by a native English speaker. Clicks were recorded as raw X and Y pixel coordinates of the point where the screen had been touched and answers were subsequently coded by matching the coordinates to the corresponding answer area among the two possible choices (left or right). Children were tested individually in a quiet room at their school. After hearing a story about

Fig. 4.1 Example stimulus experiment 3a and 3b (frequency inference), *'Oh, a cat that has a dax! Can you touch the frog that has a dax?*

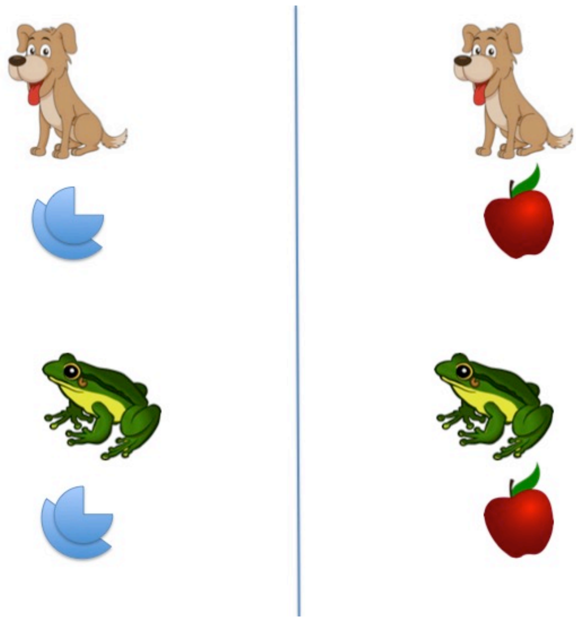


Fig. 4.2 Example stimulus experiment 3a and 3b (control/mutual exclusivity condition), *'Oh, a dog that has a dax! Can you touch the frog that has a dax?*

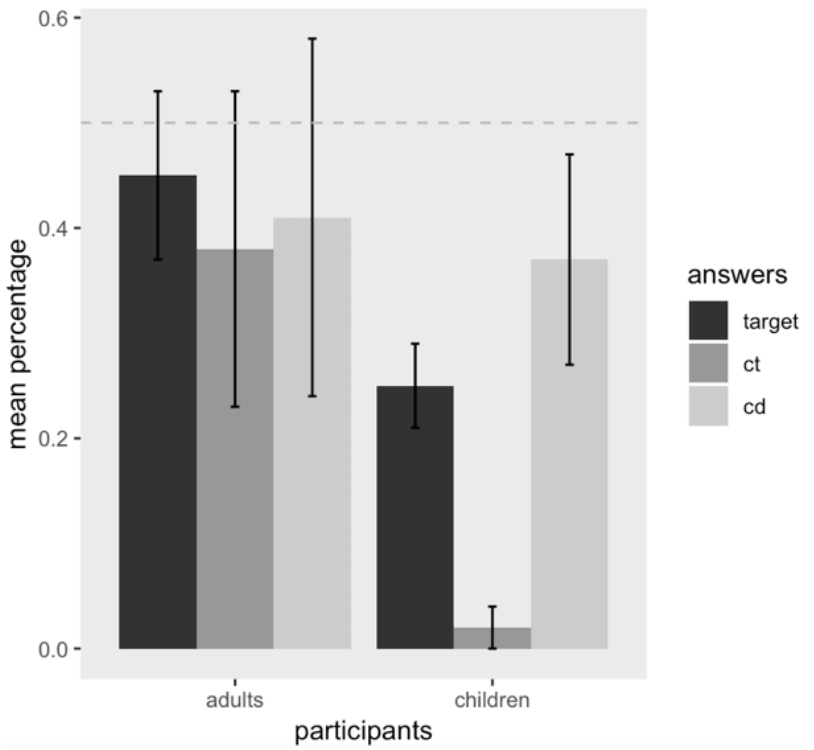


Fig. 4.3 Experiment 3a results
ct=consistent target choices (all four trials), cd=consistent distractor choices. Error bars show 95% confidence interval obtained through nonparametric bootstrapping

character Mr. Puppet meeting aliens and completing a first task, children were asked if they wanted to keep helping Mr. Puppet by finding animals who stole some of the aliens' objects. Trials were randomised, with critical and control trials interleaved. For each trial, children saw an image on the screen and heard an instruction of the type 'Oh! A kitten that has a fep! Can you show me the frog that has a fep?' using one of eight novel words (*fep, tuki, keef, razee, toma, zef, gabo* and *fid*). They then responded by touching the screen, which switched to the next trial. The experimenter sat next to the child to ensure they understood what was asked of them and completed the task appropriately, but provided only neutral feedback.

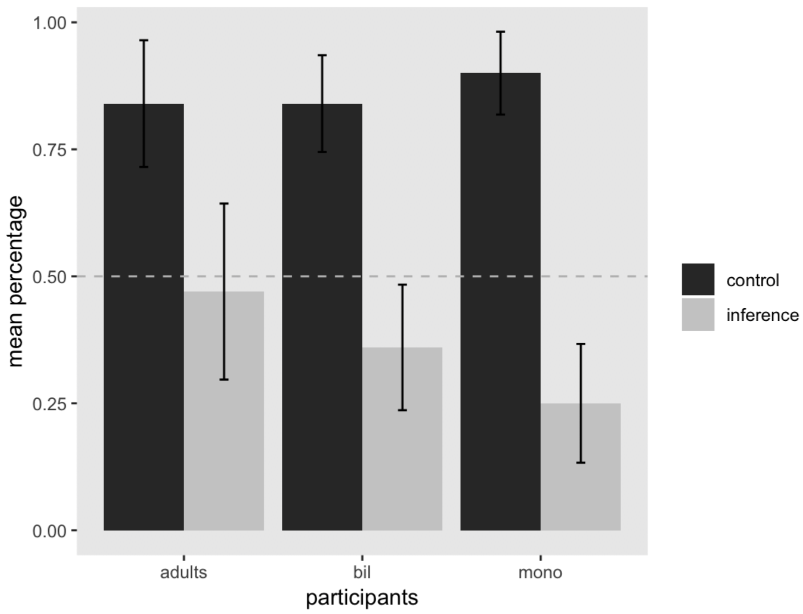


Fig. 4.4 Experiment 3a results
bil=bilinguals, mono=monolinguals

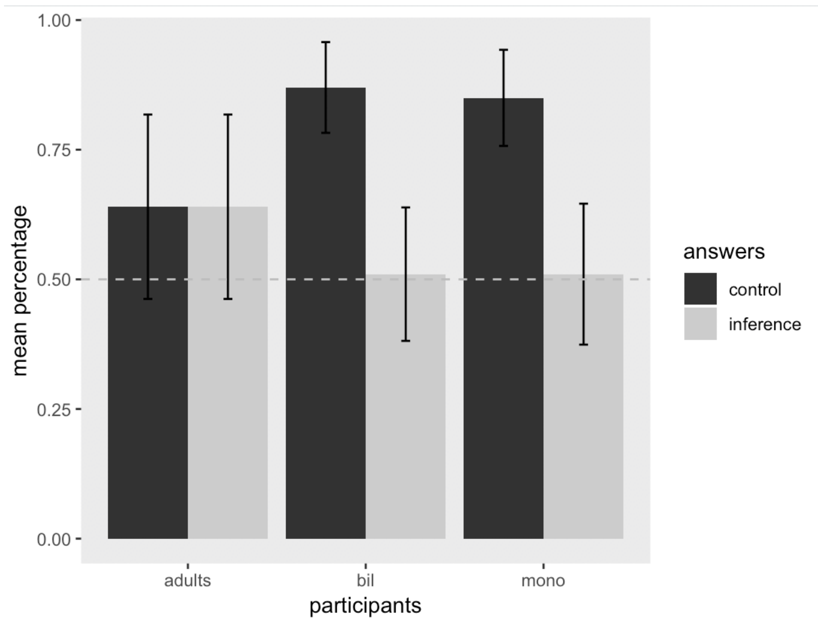


Fig. 4.5 Experiment 3b results
bil=bilinguals, mono=monolinguals

4.4.2 Results

Surprisingly, we found a distractor bias for children, with both monolinguals (m=0.25, sd=0.43) and bilinguals (m=0.36, sd=0.48) selecting distractor more often than target, whereas adults appeared to be at chance (m=0.47, sd=0.50) (see table 4.1). Interestingly, bilinguals' target choices were closer to adult scores than monolinguals'. Performance in the mutual exclusivity condition was also higher in monolinguals (m=0.90, sd=0.30) than in bilinguals (m=0.84, sd=0.37). To investigate the reliability of these effects, we again fit a logistic regression mixed model (D. Bates et al., 2014) with items and participants as random effects and starting with all design-relevant fixed effects as random slopes and removing random effects when they lead to non-convergence (Cane et al., 2017; Horowitz et al., 2018). As in experiment 1 and 2, the final model included gender, normalised BPVS scores, normalised FSM percentages, bilingual status, condition and their interaction as fixed effects, and subject intercepts as random effects. The critical frequency inference condition was used as the reference level. For adults, the model was simply specified with condition, subject and their interaction

$$Correct \sim Condition + (Condition|subject). \tag{4.1}$$

Target choices were significantly below chance in children (est=-0.64, se=0.17, z=-3.798, p=0.0001) and at chance in adults (est=-0.53, se=1.20, z=-0.44, p=0.66), with significantly higher performance in the mutual exclusivity condition both in children (est=2.25, se=0.16, z=14.45 p<0.0001) and in adults (est=2.97, se= 1.28, z=2.31, p=0.02). There was a significant effect of vocabulary scores (est=0.20, se=0.057, z=3.54 p=0.0004) and interaction with bilingualism, with monolinguals performing significantly lower than bilinguals in critical condition (est=-0.67, se=0.15, z=-4.5, p<0.0001) and significantly higher in mutual exclusivity condition (est=1.10, se=0.24,

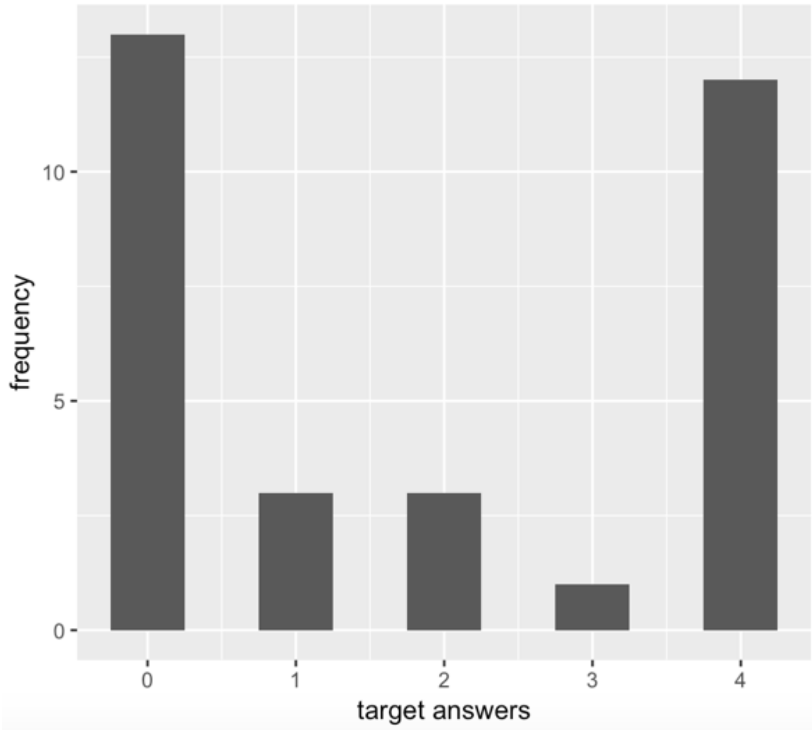


Fig. 4.6 Experiment 3a: adult's bimodal answers

z=4.565, p<0.0001). Further exploration revealed that a substantial proportion of both children and adults were consistently (i.e., in all four trials) choosing the distractor object, while a similar proportion of adults, but not children, consistently picked the target object, leading to a bimodal pattern of answers in adults verified by Hartigan's dip test (D=0.19, p <0.0001) (see figure 4.6). A graph showing the results for the adults and children can be seen in figure 4.3. The results broken down for bilingual and monolingual groups can be seen in figure 4.4.

4.4.3 Discussion

The results of Experiment 3a were contrary to model predictions and previous findings, with adults performing at chance levels and children significantly below chance. However, a closer look revealed this was mainly the result of a large percentage of both adult and children participants consistently choosing distractor over target object. What appeared as chance performance in adults was in fact due to an equivalent percentage of adults consistently choosing target object, creating a bimodal distribution of answers, whereas no alternative strategy counterbalanced distractor choices in children. One explanation for these unexpected results is that participants misinterpreted which training character (e.g., which kitten) was being referred to, despite the target character being singled out with a red circle. While this would explain random choices, it did not explain why both groups were displaying a consistent distractor bias. Furthermore, there was a clear difference between the behaviour of children and adult participants who were not consistently choosing the distractor, with non-CD children still answering in a way biased towards the distractor, whereas a large majority of non-CD adults were consistently choosing the target object. The examination of consistent patterns of answers across trials answered some of our interrogations. However, the uncertainty regarding which training character was being targeted added an additional prior parameter which was not part of the original paradigm and made any attempts at quantitative interpretations potentially circular (since the probability of choosing a training character was not entirely independent from the probability of choosing an object). For these reasons, we decided to run additional experiments. Experiment 3b replicates Experiment 3a with the targeted character made explicit in the instruction, while Experiment 4 explores whether a potential alternative quantity implicature drove the consistent distractor bias in Experiment 3a.

Bilingual performance was significantly higher (and closer to adult scores) than monolinguals'. However, given that neither group of children appeared to be performing the frequency inference, this is most likely due to bilinguals performing closer to chance/having less distractor bias than monolinguals.

While target object choices in Experiment 3a (consistent or not) could only come from choosing target training character (the only one who had the target object) and

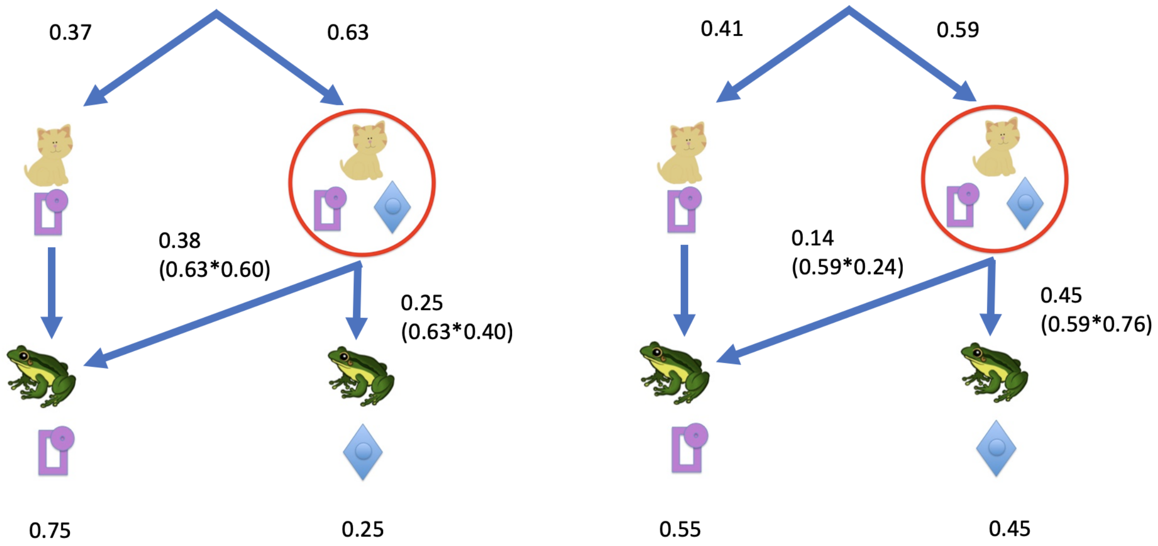


Fig. 4.7 Experiment 3a: tentative reconstruction of inference probabilities. (top: proportion choices of training kitten character after hearing *'Oh, a kitten that has a fep!* bottom: transition probabilities for test frog after *'Can you show me the frog that has a fep?)*

consistent distractor choices most likely resulted from choosing distractor training character, it is not possible to identify the source of inconsistent distractor choices (which could have resulted from choosing either training character, see Figure 4.7). In addition, the target inference might have been too difficult for children because it potentially required an extra step of probabilistic reasoning about the interlocutor (e.g., 'the blue object is probably the target since the right kitten is probably the training target since it has the most specific object' vs. 'the blue object is probably the target since the right kitten is the target and it is the most specific object it has'), which has been shown too costly to be performed sometimes even in adults (Degen et al., 2013).

Given that the percentage of consistent distractor choices in children and adults in Experiment 3a was equivalent (0.37 vs. 0.41), it seemed that the difference in performance was mainly due to the absence of consistent target pragmatic inference in children (0.02 vs. 0.38 in adults). However, since it was unclear to which extent the uncertainty regarding target training character might have driven the results (increasing

Table 4.1 Experiments 3 and 4: summary of results

Condition	Adults		Monolinguals		Bilinguals	
	mean	SD	mean	SD	mean	SD
(exp3a) mutual exclusivity	0.84	(0.36)	0.90	(0.30)	0.84	(0.37)
(exp3a) frequency inference	0.47	(0.50)	0.25	(0.43)	0.36	(0.48)
(exp3b) mutual exclusivity	0.64	(0.48)	0.85	(0.36)	0.87	(0.34)
(exp3b) frequency inference (explicit)	0.64	(0.48)	0.51	(0.50)	0.51	(0.50)
(exp4) mutual exclusivity	0.81	(0.39)	/	/	/	/
(exp4) quantity implicature	0.81	(0.39)	/	/	/	/

the difficulty by requiring two steps of probabilistic reasoning instead of one) we remove this prior uncertainty in Experiment 3b by referring explicitly to the target training character in the instruction.

4.5 Experiment 3b: frequency inference with explicit target

4.5.1 Method

Participants

Participants in the adult study were native English speakers (n=28) recruited through Prolific Academic who completed the study online for a reward of £0.37.

Children participants were 110 children aged 4 to 6 years old (m=5;5, sd=7;5) recruited in a Cambridge middle class school. The study was approved by the Cambridge Psychology Research Ethics Committee and parents in participating schools were sent information about the study along with a form allowing them to opt-out if they wished to. Since we used an opt-out policy, information regarding language status was obtained through school staff and confirmation with child, a procedure which, while unconventional, was found to be highly reliable when compared to parental forms (over

98%match) in a former study. Half the children (n=58, 30 females, 28 males, mean age=5;4, sd=7.2 months) had regular exposure to another language through one or both their parents and could demonstrate some level of fluency. The remaining children (n=52, 19 females, 33 males, mean age=5;5, sd=7.7 months) had no regular exposure to a language other than English. Two bilinguals and eight monolinguals were excluded from final analyses for fussiness, failure or inability (e.g., deafness) to complete the task. Languages were Bengali (n=8), Chinese (n=7), Polish (n=6), Afrikaans (n=4), Spanish (n=5), Urdu (n=5), French (n=3), German(n=2), Greek (n=2), Korean (n=2), Thai (n=2) Vietnamese (n=2) and Arabic, Dutch Hebrew, Hungarian, Kurdish, Portuguese, Romanian, Russian, Tamil, Telugu and Turkish (all n=1).

Receptive vocabulary

As for the first sample of children (experiments 1 and 2), participants also completed a computer version of BPVS-3 (Dunn & Dunn, 2009) implemented on the touchscreen laptop to test receptive vocabulary.

Stimuli

The stimuli were the same as in Experiment 3a.

Procedure

The procedure was identical to that of Experiment 3a, with the only difference that the instructions now made explicit mention of the target character, e.g., ‘Oh! The kitten *circled in red* has a fep! Can you show me the frog that has a fep?’

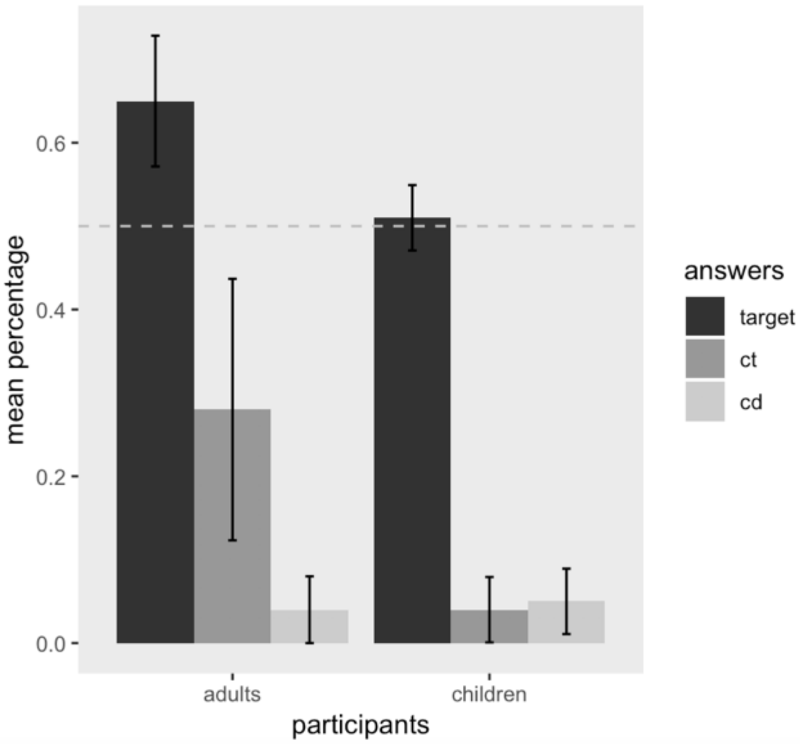


Fig. 4.8 Results Experiment 3b
ct=consistent target choices, cd=consistent distractor choices. Error bars show 95% confidence interval obtained through nonparametric bootstrapping

4.5.2 Results

Adults seemed to be performing the frequency inference (m=0.64, sd=0.48) whereas both monolinguals (m=0.51, sd=0.50) and bilinguals (m=0.51, sd=0.50) were at chance. Performance in mutual exclusivity condition was this time not higher in monolinguals (m=0.85, sd=0.36) than bilinguals (m=0.87, sd=0.34). This could be due to the sample of bilinguals having a different lexicon structure than the previous sample (cf. Byers-Heinlein & Werker, 2013). As expected, preliminary analyses revealed that the bilingual group had significantly lower vocabulary levels in English (m=65.18, sd=15.16) than the monolingual group (m=79.09, sd=17.01) (t = 14.95, df = 499.2, p<0.001). BPVS normalised scores were then entered in the regression model as a covariable. We again fit a logistic regression mixed model with items and participants as random effects and starting with all design-relevant fixed effects as random slopes and removing random effects when they lead to non-convergence. The model included gender, normalised BPVS scores, bilingual status, condition and their interaction as fixed effects, and subject intercepts as random effects. The critical frequency inference condition was used as the reference level. For adults, the model was again

$$Correct \sim Condition + (Condition|subject). \tag{4.2}$$

Adults were indeed above chance this time in selecting target (est=0.70, se=0.29, z=2.43, p=0.01) and did not significantly differ in mutual exclusivity condition (est=0.04, se=0.48, z= 0.09, p=0.93) whereas children were at chance (est=0.16, se=0.15, z=1.07, p=0.28), with significantly higher performance in mutual exclusivity condition (est=1.88, se=0.24, z=7.89 p<0.0001). There was again a significant effect of vocabulary scores (est=0.29, se=0.08, z=3.585, p=0.0003). No other effects were significant. A graph showing the results for monolinguals and bilinguals can be seen in figure 4.5, for adults and children in figure 4.8.

4.5.3 Discussion

There were this time no differences between bilinguals and monolinguals, possibly because choices appeared to be mostly random for all children.

The explicit reference to the target training character in Experiment 3b allowed us to confirm that consistent distractor choices in Experiment 3a had been driven by uncertainty regarding which training character was being referred to. However, we found results to be equivalent to the trend in Experiment 3a (when excluding consistent distractor choices) with adults, but not children, succeeding in performing the target specificity inference and selecting the target object above chance. It therefore did not seem that children's failure in deriving the target pragmatic inference was due to the increased difficulty of having to perform two steps of probabilistic reasoning, but it remained unclear what had driven consistent distractor choices in children and adults in Experiment 3a, and how this potentially related to performance in Experiment 3b. Experiment 4 was designed to explore whether this bias was the result of performing an alternative pragmatic inference which might have been easier for children to compute compared to the target inference.

The consistent distractor bias in Experiment 3a could have been due to an alternative ad hoc implicature ('a kitten that has a fep [and nothing else]', cf. Stiller et al., 2015) or to a matching strategy consisting of pairing similar descriptions (kitten/frog 'that has a fep') with similar-looking characters (kitten/frog that only has distractor object) made possible by the semantic ambiguity (both characters could be referred to as 'a kitten that has a fep' if fep was the distractor object). Arguably, this ad hoc implicature might have been easier to compute since it is based on reasoning about alternative expressions and thus benefits from multiple reinforcing cues which might make it a stronger inference ('if they had meant the kitten with two objects they would have said 'a kitten with a fep and a dax', while there is no obviously available alternative expression for the distractor training character based on the specificity inference: 'if they had meant the kitten with the most frequent object, they would have said..?')[1]. In order to test this, we ran a version of the experiment with an instruction that was pragmatically ambiguous but semantically unambiguous. The choice of the distractor character would then indicate that an alternative pragmatic inference had been performed.

4.6 Experiment 4: quantity implicature

4.6.1 Method

Participants

Participants were students of the University of Cambridge (n=82) who were recruited through email and completed the study online.

Stimuli

The stimuli were the same as that of Experiment 3a and 3b but included only the training characters, and no red circle to indicate which character was being referred to. An example stimulus can be found in figure 4.9.

[1]A crucial difference between ours and Stiller et al.'s (2015) paradigm is that the use of novel words makes it impossible to reason about an alternative expression which might have unambiguously referred to the unique object ('a kitten that has a dax' would not have been more informative than 'a kitten that has fep'; unlike 'my friend has a hat' compared to 'my friend has glasses'), thus the alternative which refers to both objects ('a fep and a dax'), despite being non-optimally informative, is the only one that could have been considered here.

Fig. 4.9 Example stimulus experiment 4,*'Oh, a kitten with a fep! Can you show me which one of the kittens is the one that has a fep?'*

Procedure

The procedure was similar to that of Experiment 3a and 3b, but the instructions used were of the type *'Oh, a kitten with a fep! Can you show me which one of the kittens is the one that has a fep?'*. The use of a reinforced definite article ('the one') ensured that this instruction was semantically unambiguous with regard to how many characters fit the description (one), thus the choice of the distractor character (with only one object) could only result from performing a pragmatic inference 'the one that has a fep [and nothing else].

4.6.2 Results

We fit a logistic regression mixed model for each of the studies with random effects for participants and condition using the lme4 package (D. Bates et al., 2014). Model specification was

$$Correct \sim Condition + (Condition|subject). \tag{4.3}$$

Participants were significantly above chance in choosing target character (b=2.66, z=5.44, p<0.0001) and performance did not significantly differ in control condition

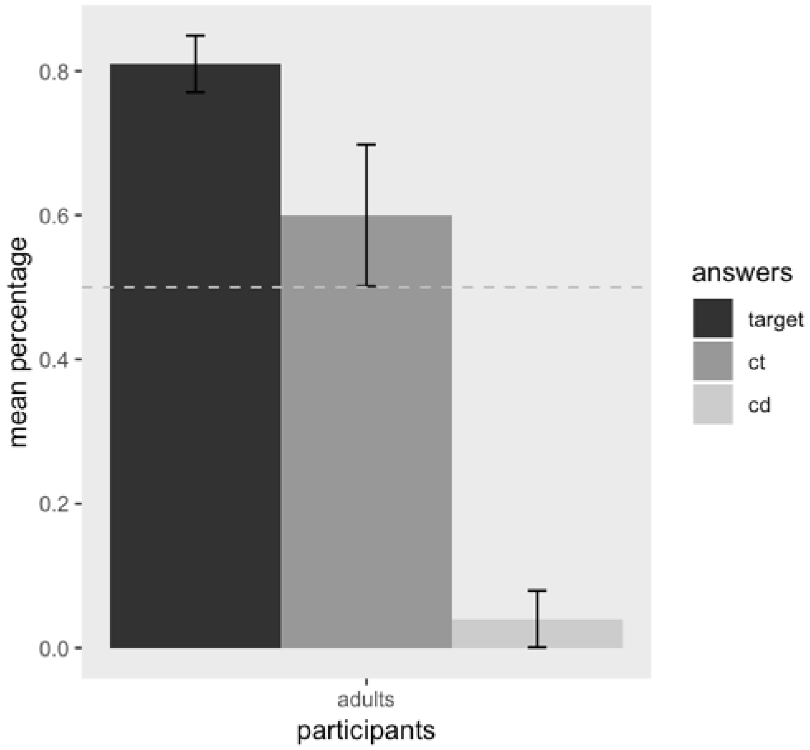


Fig. 4.10 Results Experiment 4
ct=consistent target choices, cd=consistent distractor choices. Error bars show 95% confidence interval obtained through nonparametric bootstrapping

(b=0.35, z=0.60, p=0.54). There were no consistent distractor choices. A graph showing the results can be seen in figure 4.10.

4.6.3 Discussion

There was no evidence of an alternative ad hoc implicature of the type 'the kitten that has a fep [and nothing else]' (i.e., no consistent distractor choices) in Experiment 4 and participants overwhelmingly picked the character that had two objects including the (unique) target object. We conclude that the distractor bias could not have resulted from the alternative implicature alone, since the removal of the test characters blocked distractor choices, and that this bias was more likely the result of a matching strategy consisting of pairing similar descriptions (e.g., 'the kitten/frog that has a fep') with similar-looking characters (e.g., the kitten/frog that has one object).

4.7 General Discussion

Children demonstrate a range of pragmatic abilities in language processing and acquisition and have been shown to be successful in performing pragmatic inferences based on specificity or novelty, such as those in 'ad hoc' implicatures (Stiller et al., 2015). Thus, their struggles in the derivation of scalar implicatures such as 'some' as 'some-but-not-all' have been hypothesised to be the consequence of difficulties specific to this particular type of implicature, i.e., availability of the relevant scale or alternatives, or trouble with the lexical family of quantifiers (Horowitz et al., 2018). Our failure to find successful performance of a specificity inference for word learning in children in a paradigm adapted from Frank and Goodman (2014) suggests that another explanation might be in order. A main difference between ours and Frank and Goodman's study was the simultaneous presentation of training and test items for ease of computerisation.

As the results in Experiment 3a demonstrated, this had the effect of biasing participant responses towards the distractor. Since we found no evidence in a separate experiment (Experiment 4) that this bias could be driven by an alternative ad hoc implicature alone, we hypothesise that the simultaneous set up highlighted the similarity between distractor training and test characters and made a 'distractor matching' answer a readily available and easy default option. While the explicit mention of the training character disallowed the consistent use of this strategy in Experiment 3b, we believe that the salience and availability of the matching strategy still prevented children from considering an alternative strategy and computing the pragmatic inference, resulting in chance performance. Indeed, in their study Horowitz et al. (2018) found children to struggle with both 'some' and 'none' statements and performance between the two to be correlated, whereas performance with 'all' statements was very good and approaching ceiling in 4-5 years old. As they themselves acknowledged, this could be the result of a default matching strategy which pairs statements containing a certain noun (e.g., 'some/none of the cats') with the picture that only features exemplars matching the NP (i.e., the picture with only/all cats). If this is the case, it would explain why removing 'none' trials worsened performance in their last experiment, since it gave more strength to the default matching strategy/less occasions for children to question it or consider an alternative. Similarly, this application of a default strategy across trials might have been reinforced in our study by the control condition, where the correct answer did involve pairing similar training and test characters. More generally, this strategy of simply disregarding quantifier in favour of an easier answer which only requires processing of the NP might explain poor performance in scalar implicatures since most studies use a paradigm where this option is salient (e.g., Some of the elephants have trunks; Noveck, 2001). In a study such as Papafragou and Tantalou (2004), where answers with implicatures did not match questions (*'Did you eat the oranges/the* *sandwich? -I ate some/the cheese'*) and had to be penalised (because they implied that the character did not complete the task), children performed above chance. This possibility was recently investigated by J. Sullivan et al. (2017) who indeed found that children, contrary to adults, withheld prizes when the surface form of Puppy's reported action did not match the form of the request even when the result entailed the request (e.g. *'Did you colour some of the stars?' 'I coloured all'*). Crucially, children did give prizes when the same result was shown in a picture rather than labelled with a contrasting linguistic form, which suggests that these results were not due to them first computing a scalar implicature but rather to the influence of surface linguistic forms. An exception is a paradigm such as the one in Katsos and Bishop (2011), where no binary answer (right/wrong judgment or picture choice) is required and 'matched' pairings are being explicitly compared in a qualitative way (using a graded reward system), which draws children's attention towards the relative appropriateness of the default strategy (i.e., is pairing 'some of the carrots' with a picture containing all the carrots as good as pairing it with 'all of the carrots'?), resulting in increased pragmatic sensitivity and better performance.

Why do adults succeed in performing these inferences? As Experiment 3a showed, adults like children are sensitive to the presence of a salient 'easier' matching answer, however the availability of this strategy does not prevent them from considering an alternative pragmatic answer. Furthermore, when this matching strategy is explicitly disallowed, the adults' world knowledge/semantic filter will efficiently block it in favour of semantically appropriate answers, whereas it will still exert some (albeit reduced) bias on children (just like children's answers on 'some' and 'none' trials are not quite on par with 'all' trials, but still either at or above chance in Horowitz et al.). The tendency to group things on the basis of similarity is most likely a cognitive bias which is part of the set of domain-general learning tools children are readily equipped with: categorisation will primarily proceed in a similarity-based fashion, with language-specific exceptions starting to appear as late as 10 years old (Storms et al., 2015). Pragmatic inference, on the other hand, has to be learned and requires more computational effort. In the case of scalar implicatures and lexical items, adults' more extensive linguistic experience probably means that processing quantifiers requires less effort as it has become more default, and thus more likely to overcome a salient answer based on similarity. However, our study seems to indicate that this adult advantage persists even in cases where the inferences to be performed are not conventional or generalised, which could result from more experience with deriving pragmatic inferences in general or from a better ability to inhibit a salient answer to consider an alternative strategy. While Horowitz et al. (2018) failed to find a correlation between performance on an inhibitory control task (the Dimensional Change Card Sort or DCCS, Zelazo, 2006) and on some/none trials, it is worth mentioning that this task requires only to inhibit a 'learned' response (i.e., answering based on shape or colour) which participants have been applying for a few trials, rather than a 'naturally salient' answer based on matching or similarity. Given this difference, a Stroop task or a task based on inhibiting a NP-picture matching (such as the Day/Night task) where linguistic information is explicitly pitched against a naturally salient answer might be more appropriate and future research should focus on investigating a potential correlation with these measures. A recent study showed no differences between a group of participants with ASD (Autism Spectrum Disorder) and neurotypical controls on either deriving scalar implicatures or performing epistemic reasoning when explicitly asked, however the neurotypical group proved to be better at cancelling the implicature when it was unwarranted (Hochstein et al., 2018). This suggests both that this type of implicature is at least partially defaultly or grammatically derived, and that the difference between the two groups is not one of general theory of mind ability but rather flexibility in making spontaneous 'online' use of their epistemic reasoning when required.

While children's success in choosing target object above chance in Frank and Goodman's (2014) study is most likely due to the absence of a salient matching strategy, another possibility has to be considered. The assumption in Frank and Goodman's model is that feature choices are proportional to the relative frequency of each feature regardless of whether the feature is unique or not. However, even in adults' responses, the relationship between frequency ratio and choices appears to be more logarithmic than linear in nature, with uniqueness of one feature having a considerable effect on participant choices (0.57 vs. 0.67) whereas when the feature is unique the exact ratio (1/2 or 1/3) seems to matter little (0.70 vs 0.67). This suggests that there might be a more qualitative difference between cases where the target feature is unique and those where it is not. One explanation could be that the reference to the target character is interpreted contrastively (i.e., *'This is a 'daxy' dinosaur/dinosaur with a dax* [and these others are not]'). While Frank and Goodman (2014) do attempt to control for this possibility by using neutral prosody and a control condition to test whether uniqueness of the feature alone could have driven performance, this explanation could only be completely ruled out by testing children on trials where the target feature is not unique and/or the distractor training character possesses as many features as the target training character[2]. The simultaneous presentation of training and test characters in our study had the effect of eliminating the 'uniqueness' of the target feature when the full array (all four characters) was taken into consideration and 'diluting' the feature's informativeness or frequency ratio from 1/2 to 2/3 (or 0.67 to 0.60 according to model predictions), which might have contributed to children's

[2]In the condition designed by Frank and Goodman to control for this uniqueness/contrastive effect, children are asked 'can you find another one?' after being told in training that 'this is a dinosaur with a dax', which does result in chance performance, but this could have simply been interpreted by the children as 'can you find another dinosaur'.

failures and lower rates of consistent choices even in adults. Children in the original study might have been performing a simpler exclusion inference (cf. J. Sullivan et al., 2017), which would perhaps also explain why their performances were above model predictions even with rationality parameter equal to 1. Informativeness reasoning in children might thus proceed in a more 'all or nothing' than proportional fashion, with a qualitatively different strategy (i.e., exclusion inference) in simpler cases.

When the distractor choice was not explicitly disallowed, bilinguals appeared to have less of a distractor bias than monolinguals. This could be because they had a weaker similarity bias, or because they were more attentive to all cues, especially ostensive ones, and thus potentially more likely to have attended to or have been made hesitant by the presence of the red circle, while still not being able to perform the frequency inference. This could be related to an ability to inhibit irrelevant information displayed by bilingual children in certain tasks , especially in favour of social cues (cf. Chapter 2), i.e., bilinguals could be more flexible and more able to momentarily inhibit a first response to examine all possible cues or responses available (particularly ostensive cues such as the red circle).

In conclusion, we presented several experiments designed to investigate whether a bilingual advantage in word learning using pragmatics would be more apparent when only a minimal amount of pragmatic information ('one cue') was provided. We hypothesised that children might be able to perform such word learning using only one cue in a case where the inference was made stronger because it arose from potential *underinformativeness*, rather than *overinformativeness*, on the part of the speaker. However, contrary to our expectations and to previous results, we did not find children to be able to succeed in this task. We then hypothesise that children might have been performing a simpler type of exclusion inference in previous studies and that their lack of experience with considering pragmatic information in balance with other, more salient sources of information might have led them to answer following an easier 'default' path based on similarity in our experiments. Given that bilinguals, as in other, previous studies (cf., Brojde et al., 2012) did not display as strong a similarity bias towards a 'default' matching answer, we hypothesise that they might exhibit a better ability to flexibly consider and balance a diversity of alternative sources of information and/or particularly when this involves taking into account pragmatic cues.

Chapter 5

Ostension and extension

That's a beautiful speech, but nobody's listening. Let's go.

— Alfred Jarry

5.1 Ostensive teaching

If during a party you start yawning or looking at your watch while exchanging insistent glances with your wife across the room, she will probably understand your intention to communicate that you are bored and want to leave. If, on the other hand, she witnesses the same yawn when coming into the living room where you sit reading, she might not draw the same conclusions.

Communication is generally thought of as *ostensive*, in that it involves not only the transmission of information (an *informative* intention) but also the recognition of this intention by the hearer (a *communicative* intention) (Sperber & Wilson, 1986) which contributes to the derivation of the intended meaning. The main purpose of this act of communication will often be to transmit episodic information about a specific event or entity or to get the hearer to identify a referent, and/or perform some action on the world. For example, in the experiments presented in chapters 3 and 4, the focus was (seemingly) to get the hearer to select the correct referent to achieve some external purpose that the speaker had in mind (e.g., help Mr. Puppet). In this context of a one-instance 'fast mapping', the learning of the novel word's meaning appears (at least from the perspective of the hearer and what they have evidence to assume are the intentions of the speaker) to be incidental or collateral. This type of communication is, in fact, not limited to humans, as other species such as bees, birds and monkeys have been demonstrated to be able to communicate episodic 'here-and-now' information and to give injunctions (e.g., food location and warnings) (King, 1994).

There is, however, another type of communication which appears to be specific to our species and might well be what most crucially distinguishes it from others: the ability to perform a one-instance transmission of *referential* (e.g., 'This is a dog') or *generic* (e.g., 'Dogs bark') knowledge, effectively shortcutting the slower, basic cognitive process of induction (e.g., cross-situationally learn that 'dog' can be heard and mapped to this particular animal more often that any other referent, or witness many instances of dogs barking). According to Csibra and Gergely (2009), humans have developed a specific system for the communication of this type of generic knowledge, which they refer to as *natural pedagogy*. Their hypothesis states that there is a limited set of ostensive signals, such as child-directed speech, eye gaze and pointing, which can be used for the purposes of teaching referential knowledge. The particular sensitivity to these specific signals then allows the learning of knowledge which would be difficult to acquire by relying only on perceptual information. This bias would act as a catalyst which ensured that information was quickly encoded as generalisable. However, other works have suggested that any attention-grabbing cues, even those which do not have a conventional communicative purpose (e.g., shivering), could be used to successfully prepare children for the conveyance of referential information (Szufnarowska, Rohlfing, Fawcett, & Gredebäck, 2014). In this view, there would be no dedicated ostensive system and the mechanisms for monitoring social cues would be purely attentional, e.g., infants would learn to follow eye gaze because this allowed them to better predict where interesting visual events were likely to occur (Triesch, Teuscher, Deák, & Carlson, 2006).

Putting aside the debate of whether there is a particular set of ostensive signals dedicated and most efficient for teaching referential knowledge, our purpose was to investigate the use of ostensive teaching for word learning with signals that could *not* be used directly (as pointing and eye gaze could be) for finding the referent of a novel word. The reason is that we wanted to examine the ability of monolingual and bilingual children to perform true pragmatic reasoning about speaker's intentions to learn the meaning of a new word. As mentioned in chapter 3, there are at least three hypotheses which could explain the previously found advantage of bilinguals versus monolinguals in paradigms using 'socio-pragmatic' cues for word learning. The first hypothesis is that, given that most of these tasks could be succeeded at by ignoring a misleading cue to focus on the social cue, these results could simply have been driven by a better inhibitory ability. The second hypothesis is that bilinguals are particularly able to focus on these specific types of social cues or signals compared to other types of cues. Finally, the third hypothesis states that bilingual children are particularly apt at using pragmatic reasoning to learn new words, that is to say at reasoning about the intentions of the speaker, or *why* they provided this particular cue. If we return to the dichotomy between *communicative* and *informative* intention mentioned in the first section, it is possible that bilinguals were able to understand the communicative and ostensive intention behind the point or gaze (and understand that they had to ignore other cues) and then simply better able to follow the point or gaze and ignore the irrelevant cues, without reasoning about *how* the evidence had been provided by the speaker.

In chapter 3, we examined whether monolinguals and bilingually exposed children were able to 'pick up' on the relevance of a pragmatic cue (adjectival modifier and prosodic stress) to understand a particular communicative and informative intention without an ostensive intention to teach a novel word. Here, on the other hand, the ostensive teaching paradigm allows us to make this teaching intention clear to all participants by using types of settings which children are familiar with (i.e., presenting them with exemplars of objects). Thus the difficulty in such paradigms resides in deriving the intentions underlying the particular way in which the cues were presented by the teacher. Since we wanted to avoid ceiling effects in order to obtain enough variability between groups to compare performance, we focused on paradigms in which the correct answer was not the 'default' one as predicted by word learning biases: extending a word to the subordinate level (rather than the basic level) and to a novel action (rather than a novel object). It is worth noting that, while both of these experiments involved an ostensive teaching intention and eaching setting (prompting the awareness that a new word was being taught), as well as ignoring a 'default' response, (as in previous studies by Yow and colleagues), success (or ignoring the salient/default response) could once again not be achieved by simply following eye gaze or point, but had to involve reasoning about the speaker's intentions or *why* the cues had been presented in this way. In the next sections we briefly discuss research investigating the use of ostensive teaching for learning subordinate categories and action words.

5.2 Learning categories using exemplars

Imagine you are learning words from a foreign language speaker. You witness them search a bag which contains different types of apples (say, golden, braeburn and granny smith), take out a golden apple, present it to you and say *cabo*, then take out another golden, present it to you and say *cabo* again. In another scenario (featuring a less sturdy bag), the same golden apple falls out of the bag, they pick it up and present it to you saying *cabo*. Then, when another golden rolls out of the bag, you readily point at it and enthusiastically utter *cabo* while they nod in approval.

At which hierarchical level should you extend the novel word? Does *cabo* mean apple or golden? If you are sensitive to the source or sampling process of the exemplars, you are probably more likely to attribute the meaning of 'golden apple' to the unknown word in the first scenario. This would be extending the word to the *subordinate level*, whereas 'apple' would constitute the *basic-level* in this case ('fruit' being the *superordinate level*). This is indeed what research shows. While being presented with one exemplar generally leads to a basic-level bias (extending the word to all subordinate categories, e.g., all types of apples rather than to golden apples only, cf. Rosch et al., 1976; Waxman & Kosowski, 1990), being given several exemplars from a specific subordinate category drastically reduces basic-level choices of extension in favour of subordinate ones, as per the 'suspicious coincidence' effect (Xu & Tenenbaum, 2007b).

In their study, Xu and Tenenbaum (2007b) presented children and adults with either one or three exemplars from diverse category levels (e.g., one dalmatian dog, three dalmatian dogs, one dalmatian, one labrador and one hush puppy, etc.) which were given a new name by a puppet who was 'learning a new language' (e.g., *'This is a blick'*) then asked to select among a large set of matching options from subordinate, basic and superordinate levels (e.g., *'Can you pick out the other blicks?'*). They found that being given one exemplar led to a graded pattern of choices based on similarity, with a bias for basic-level choices. With one exemplar as evidence for the new category name, adult participants were selecting subordinate matches almost all of the time (97%) and basic level matches most of the time (76%). On the other hand, when given several exemplars, participants were generalising in a more all-or-nothing manner, selecting all matches from the most specific or narrowest hypothesis (e.g., selecting all dalmatians after having seen three dalmatians) and almost no matches from broader categories (e.g., no other dogs), as per the *Size Principle* and 'suspicious coincidence' effect (Tenenbaum & Griffiths, 2001). Indeed, the exemplars have a probability of being sampled which is proportional to the size of each category or nested sets of entities, e.g., all animals, all dogs, all dalmatians, etc. Thus, given that there are by definition more dogs than dalmatians in the world, the probability of a sample including only exemplars from the subordinate category of dalmatians is low ('suspicious') and decreases exponentially with each new exemplar, unless the label only extends to this particular subordinate category. Xu and Tenenbaum (2007b) model this 'one-versus-three' difference for word learning inferences using a Bayesian model where prior probabilities for each category are approximated through branch height in a taxonomic tree (i.e., the higher the hierarchical level, the more entities it contains) and likelihoods are the combined probabilities for each additional exemplar of having been sampled from this specific category. In a non-biased Bayesian model, the prior probabilities are thus considered equal for all category levels, i.e., (given that they are nested categories): 100% for subordinate, 50% for basic level, 25% for superordinate. However, the authors found that they needed to modify the prior to include a basic level bias in order to accurately model adult participants' performances. This bias towards extending a novel word at an intermediate 'basic' taxonomic level was first described in the work of Rosch et al. (1976). The reason for this preference is hypothesised to be that some intermediate categories are considered 'basic' from a cognitive perspective. These categories benefit from a privileged psychological status by virtue of having a number of common attributes which is considered optimal in terms of utility, that is they have sufficient predictive power and a sufficient level of abstraction. More specifically, basic level category members are described as having similar shapes and motor programs, which means for most predictive purposes they are interchangeable (e.g., eating a golden or braeburn apple has roughly the same effect and a dalmatian or labrador met on the street will behave in a similar way). As for other word learning constraints, the question of whether this basic level bias is present from an early age and contributes to reducing the hypothesis space of referents from the very beginning is debated (cf. Callanan, Repp, McCarthy, & Latzke, 1994).

In addition to this, learners are more likely to extend a word in a narrow way (e.g., at the subordinate level) if the multiple examples have been provided through *strong* sampling (i.e., by a knowledgeable speaker who is sampling the true meaning distribution) rather than *weak* sampling (i.e., randomly among the full hypothesis space, or conservatively by an ignorant learner hoping for a reward) (Xu & Tenenbaum, 2007a).

5.3 Learning words for actions

Words for actions are generally considered more difficult to learn than words for objects. Some have argued that this is due to conceptual reasons, because nouns, unlike verbs, refer to units which are distinguished perceptually as well as consistent across time and context (Gentner, 1982). As we have seen in the introductory chapter, there might be a bias or default assumption which drives young children to prefer mapping a new word onto a new unnamed whole object rather than onto any alternative referent, including familiar objects, parts, properties and actions (Markman, 1994). This could lead to the apparent noun bias witnessed in early vocabularies across many languages such as English (Jackson-Maldonado, Thal, Marchman, Bates, & Gutierrez-Clellen, 1993), French (Poulin-Dubois, Graham, & Sippola, 1995), Italian (Caselli et al., 1995) and Hebrew (Dromi, 1999). Few studies have in fact examined the preference for whole objects over new actions in morphosyntactically neutral contexts, probably partially because of the difficulty in creating such contexts for languages which have rich morphologies. The few studies which did directly investigate this preference did find a bias towards object mappings in children (Imai et al., 2008; Katerelos, Poulin-Dubois, & Oshima-Takane, 2011; Woodward, 1993). However, this preference for object mappings appears to be overridden through the use of socio-pragmatic cues (Tomasello & Akhtar, 1995). In Tomasello and Akhtar's (1995) study, two year olds were presented with a novel word, uttered while performing a novel action with a novel object. In one condition, the previous context was designed to highlight the novelty of the action relative to the common ground between child and experimenter (several actions/manipulations were performed on the object for a length of two minutes before performing the target action while uttering the novel word). In the other condition, it was the novelty of the object which was highlighted by performing the same novel action on several different objects before performing it with target object while uttering the novel word. Results seemed to show that the two year olds were able to reliably map the novel word to the novel action in the condition where the novelty of the action had been highlighted through pragmatic cues.

Studies which more specifically investigated the acquisition of verbs found young children to struggle in some experimental paradigms even when verbs were clearly flagged as such with morphosyntactic markers (Golinkoff, Jacquet, Hirsh-Pasek, & Nandakumar, 1996; Imai, Haryu, Okada, Lianjing, & Shigematsu, 2006; Imai et al., 2008; Kersten & Smith, 2002; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008), whereas in other paradigms even two year olds could succeed (Gertner, Fisher, & Eisengart, 2006; Noble, Rowland, & Pine, 2011; Oshima-Takane, Ariyama, Kobayashi, Katerelos, & Poulin-Dubois, 2011). The main difference between these paradigms seems to be the type of testing. In Naigles (1990) for example, two year olds were taught a novel verb with a familiar object and object name (e.g., *'The duck is gorping the bunny'*) and simply asked to pick between two different actions at test time. In contrast, in Imai et al. (2008), three and five year olds were exposed to a video of an agent performing a novel action with a novel object while uttering a novel word and asked at test time to choose between a video showing the target action performed with another object (Action Same) or a video of another distractor action performed on the original novel object (Object Same). In this latter case, the child was required not only to extend and generalise the verb to another object at test but also to ignore an answer involving the original object used to demonstrate the taught action.

This shows that attentional aspects of a task are probably the main drivers of performance for young children, who will only be able to succeed if processing load is reduced to a minimum and there are no distracting features, which constitutes a methodological issue in experimental designs where distractors are often needed. Accordingly, length of exposure seems to override other effects, such as semantic properties, in determining the difficulty for learning a given verb type. For example, Abbot-Smith, Imai, Durrant, and Nurmsoo (2017) conducted a study examining whether verbs for resultative actions (leading to a change in state) were more easily learned than verbs for durative actions. Contrary to expectations, the reverse was found to be true, which they hypothesised could be due to the visual brevity of punctual actions making them more difficult to encode, especially for young children who have slower visual processing speeds than adults (Liss & Haith, 1970).

The goal of the following experiments was to examine whether young bilingual and monolingual children showed differences in the way they use ostensive teaching to map and extend a novel word when pragmatic reasoning (i.e., reasoning about the intentions underlying the cues provided by the 'teacher') was required to select the target referent rather than responding according to a default bias or preference. We wished to examine 1) whether children could perform a pragmatic inference on the basis of a single cue in an ostensive teaching setting where the didactic/communicative intention was clearer 2) if bilingual children could, in such a setting, use the pragmatic reasoning ability demonstrated in Experiment 1 to ignore a default response and correctly extend the meaning of a novel word, 3) whether a better matched (in age and socio-economic background) sample of bilingual and monolingual children would this time display not just a relative but a numerical advantage and 4) whether the relative bilingual advantage found in Experiment 1 would apply not just to fast mapping but also to extending a novel word.

5.4 Experiments 5 and 6

Children were able to perform fast mapping for referent disambiguation using pragmatic inference when two cues (modifier and stress) were available but not when only one cue (modifier) was available as in experiment 1, even in the context of underinformativeness as in experiment 3. In addition, bilingual children had a relative advantage over monolingual children in performing contrastive inference for fast mapping in experiment 1. Given these results, we wanted to investigate the following: 1) are children able to perform pragmatic word learning using only one cue in the more supportive context of ostensive teaching? 2) In this more supportive context, do bilinguals display an advantage, even in tasks where extension of the newly learned word is required?

5.4.1 Experiment 5a: subordinate category teaching

The purpose of this experiment was to examine how bilingual and monolingual children performed word learning of a subordinate category taught using ostensive teaching. In the critical condition of this task, a 'teacher' character (ie., 'Mr Puppet') presented the children with two exemplars of a novel object at training time and uttered a prompt using a novel word (eg., *'Oh, look, Mr Puppet wants to show you a dax. And look, Mr. Puppet wants to show you another dax!*). In contrast, in the control condition the exemplars were not provided by a teacher but randomly (eg., *'Oh, look, a dax fell out of the bag. And look, another dax fell out of the bag!*). At test time, the children were presented with six objects, some of which were exactly identical to the original objects ('subordinate' matches) while others were similar but differed in colour or orientation ('basic-level' matches). The children had to pick which objects they thought were other exemplars of the new word (eg., *'Can you touch the other daxes?*. The inference in this task relies on the assumption that exemplars provided by a knowledgeable teacher are more informative than randomly provided exemplars, ie., the children should be more likely to extend the novel word to the subordinate matches in the critical, teacher condition than in the control condition. In addition, given that bilinguals displayed a relative advantage in pragmatic inference over monolinguals in our first experiment, we hypothesise that they would display a similar advantage in this task.

The task was adapted from Xu and Tenenbaum (2007a) and implemented as a computerised task.

Participants

Participants were the same 270 children as in experiments 1, 2 and 3a (cf. Chapters 3 and 4). 17 children (7 bilinguals and 10 monolinguals) were excluded from final analyses for failing to complete the task (n=12) or loss of data (n=5). Experimental design was between-subject and participants were randomly assigned to either random or teacher condition.

Stimuli

Stimuli were four sets of novel objects for the target 'basic level' categories characterised by shape, each with three subordinate categories characterised by colour and orientation (12 in total). In addition, there were two distractor objects from another category for each target set (8 in total). Four novel labels were used (*skol, liput, murbil* and *dax*). A pilot study helped adjust the characteristics of the subordinate categories so that they could be recognisable by children as one category but would not lead to ceiling effects (i.e., only subordinate level answers), as the variability was needed to compare the two language groups.An example stimulus for the teacher condition can be found in figure 5.1, and for the random condition in figure 5.2. An example for the test phase can be found in figure 5.3. A full list of stimuli can be found in Appendix 4.

Procedure

Stimuli were implemented and presented on a touchscreen laptop using Superlab version 5.0 (Cedrus, San Pedro, CA). Clicks were recorded as raw X and Y pixel coordinates of the point where the screen had been touched and answers were subsequently coded by matching the coordinates to the corresponding answer area among the six possible choices on the picture. Children were tested in a quiet room at their school. They were introduced through the computer to Mr. Puppet, who had recently made some alien friends. They were then told that Mr. Puppet had acquired some new alien objects and wanted to show these objects to them (in the teacher condition) or that the aliens had left a bag with new objects (in the random/control condition).

Fig. 5.1 Example stimulus experiment 5a (teacher condition), *'Mr Puppet wants to show you a dax. And look, Mr Puppet wants to show you another dax!'*

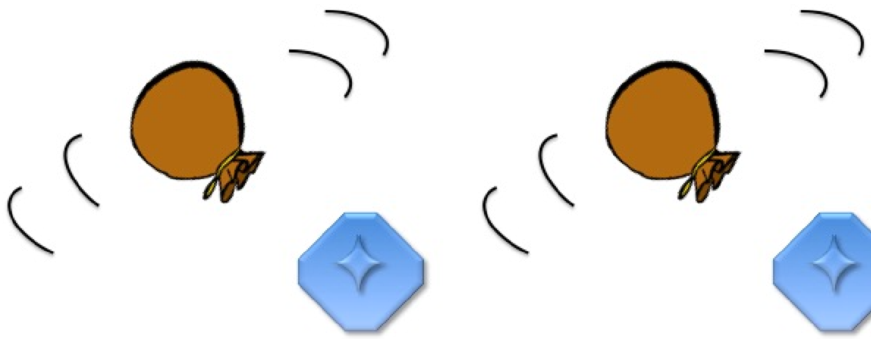


Fig. 5.2 Example stimulus experiment 5a (random condition), *'Look, a dax fell out of the bag. And look, another dax fell out of the bag!*

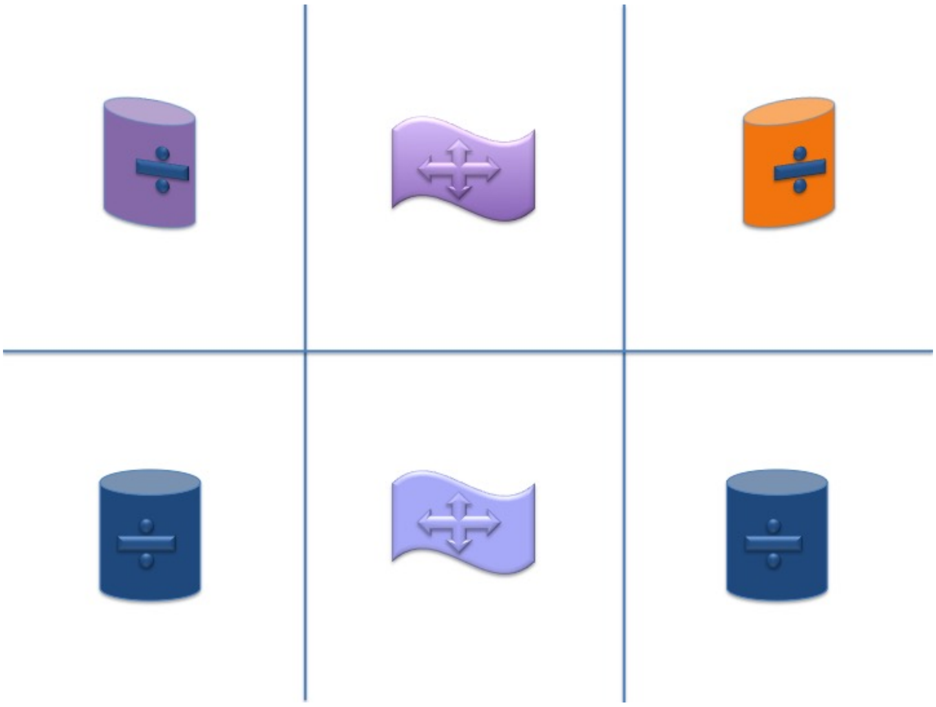


Fig. 5.3 Example stimulus experiment 5a (test phase), *'Can you show me the other daxes?'*

In the training phase of the teacher condition, children saw an image of Mr. Puppet along with an object from the target subordinate category and heard a prompt of the type *'Mr. Puppet wants to show you a skol. Look, this is a skol!'*, the screen then switched and another exemplar of the same subordinate category then appeared while the children heard another prompt of the type *'And, look, Mr. Puppet wants to show you another skol!'* (see Figure 5.1). In the training phase of the random/control condition they first saw a picture of a bag and an exemplar of the target subordinate category and heard a prompt of the type *'Oh, look what fell out of the bag, it's a liput! A liput fell out of the bag!'*, then the screen switched and they saw the bag again along with another exemplar of the same subordinate category (see Figure 5.2) and heard a similar prompt of the type *'And look, another liput fell out of the bag!'*

In the test condition the children were presented with a grid containing six objects: two exemplars of the target subordinate category(e.g., bottom left and bottom right in figure 5.3), two exemplars of other subordinate categories(e.g., top left and top right) , and two distractors from another object category (e.g., middle top and bottom). They were prompted with instruction of type *'Can you touch the other skols?'*. Subordinate answers would involve picking only the target subtypes, whereas if the novel word was interpreted at the basic category level all subtypes of the target category sould be selected. The experimenter would switch to the next trial when they got a clear signal by which the child manifested being done, either by looking at the experimenter or by having placed their hands back at their sides and appearing to be waiting for the next instruction. This procedure was used since more prompts from the experimenter could have been misinterpreted as meaning that the right/complete answer was the basic level category (*'Can you see another one/Can you see more?'*, etc.) and ensured minimal bias from feedback.

5.4.2 Experiment 6a: action word teaching

The purpose of this experiment was to investigate another type of ostensive teaching where a default bias has to be overcome using reasoning about informativeness. In this task, children were presented with a video of a 'teacher' manipulating objects in different ways. In the control condition, the 'teacher' would perform a novel action with a novel object while uttering a novel word with no morphological markers indicating whether the part of speech was verb or noun (eg.,'Dax'). In the critical condition, the 'teacher' would perform a first action with the novel object before performing the target action while uttering the novel word. The children were then presented with two side-by-side videos of the same teacher, one where they performed the target action with a different object and one where they performed a different action with the same object. They were asked to choose between the two (eg., *'Can you show me dax?'*, indicating in this way whether they believed that the word referred to the action or the object . The inference was based on the assumption that the fact that the teacher had used the novel word in conjunction with the second action was informative, ie., that the most novel referent was the most likely(the novel target action was more novel when they word was uttered than the object which had already been seen and manipulated). Given bilingual and monolingual performance in experiment 1, we predicted that bilinguals would outperform monolinguals in using pragmatic inference for word learning.

The task was inspired from Tomasello and Akhtar (1995) but we used a neutral control condition instead of a object-enhancing condition to obtain a reliable measure of object bias. Testing method was adapted from Imai et al. (2008).

Participants

Participants were the same 270 children as in Experiment 6a. 12 children (8 bilinguals and 4 monolinguals) were excluded from analyses for fussiness (n=6) or loss of data (n=6). Experimental design was within-subject.

Stimuli

Stimuli for training were 8 videos picturing a young woman (not the experimenter) performing a novel action with a novel object while looking at the camera. In the four videos of the control neutral condition the woman would perform the novel action with the novel object once while uttering a novel word. In the four videos of the action-enhancing condition two actions would be performed on the object: first a distractor action and then, while uttering the novel word, the target action (see Figure 3). While this meant that the videos differed in length, pilot studies showed that video length tended in general to increase the object bias (by means of longer exposure) regardless of condition, which meant a measure of action answers in this condition was

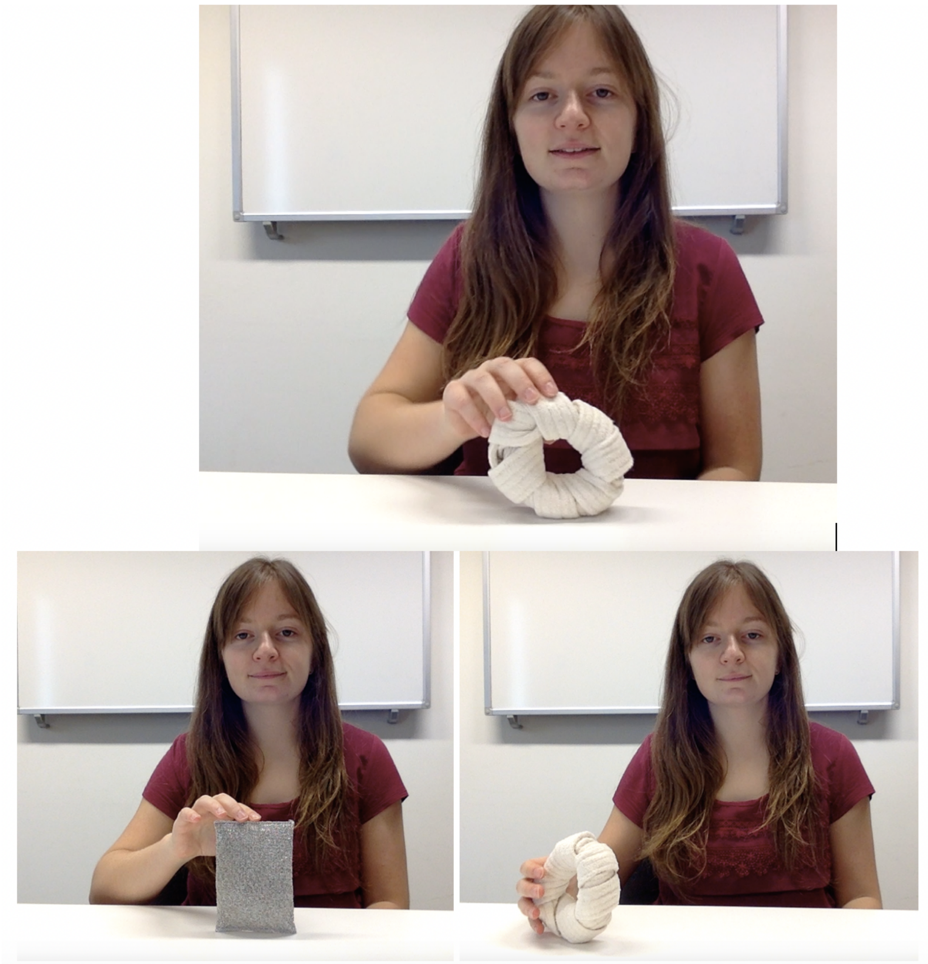


Fig. 5.4 Example training (*top, 'Dax'*) and test (*bottom, 'Can you show me dax?'*) stimuli experiment 6a

conservative. An example stimulus can be found in figure 5.4. A full list of stimuli (novel objects used, the videos can be provided if needed) can be found in Appendix 5.

In addition, there were 4 sets of two test videos where the woman would perform the target action with a different object (Action Same or AS) and a distractor action (different from the training) on the target object (Object Same or OS). Four novel words were used (*pef, dalp, squel* and *wige*) The objects were unusual or assembled from parts of other objects so as to be unrecognisable to children (see Figure 5.3). A full list of stimuli can be found in Appendix 5.

Procedure

Stimuli were implemented and presented on the same touchscreen laptop using Superlab version 5.0 (Cedrus, San Pedro, CA). Clicks were recorded as raw X and Y pixel coordinates of the point where the screen had been touched and answers were subsequently coded by matching the coordinates to the corresponding answer. The children were told that they were going to watch Mr. Puppet's friend play with some new toys he had received from the visiting aliens. They first completed two warm up trials where they were asked *'Can you show me balloon?'* and had to choose between a video of the experimenter manipulating a balloon and another video of her manipulating a pencil, then children were asked *'Can you show me clap?'* and had to choose between a video of the experimenter clapping and another of her swaying from left to right. This was done to ensure that children knew that both object and action choices were possible. As in Tomasello and Akhtar (1995), no articles or tenses were used as they would act as morphological markers giving away the word's part of speech. For each of the four trials, the child would first watch on the screen the training video (neutral or object-enhancing depending on the condition they had been assigned), then the screen would switch to the test videos (Action Same and Object Same), which were placed side by side and played on repeat (the position of AS and OS videos left or right was counterbalanced). They were asked a similar question as in the warm-up, e.g., *'Can you show me dalp?'*. The child then made their answer by touching one of the videos and the screen switched to the next trial.

5.4.3 Results

Answers in the subordinate category teaching were coded as 'basic level' if all four objects from the target category had been selected and as 'subordinate level' if only the two objects from the target subordinate category had been selected. Answers

Table 5.1 Experiment 5a and 6a: summary of results

Condition	Monolinguals		Bilinguals	
	mean	SD	mean	SD
(exp5a) subordinate category (control)	0.23	(0.43)	0.29	(0.45)
(exp5a) subordinate category (teacher condition)	0.28	(0.45)	0.38	(0.49)
(exp6a) action verb (control)	0.44	(0.50)	0.44	(0.49)
(exp6a) action verb (teacher condition)	0.44	(0.50)	0.42	(0.49)

which did not fit either pattern of response (24%) were excluded from analyses. Target responses in Experiment 5a were higher in teacher than random condition, but not in Experiment 6a where action responses were equal in both conditions (see final results in Table 5.1). A preliminary t-test indicated a significant effect of condition in Experiment 5a (t=2.52, df=815.56, p=0.01). To investigate the reliability of these results when taking into account subject and item effects, we again ran a logistic regression mixed model (D. Bates et al., 2014). Model included task, gender (shown to have an effect on pragmatics, cf. Stiller et al., 2015), bilingual status, condition and their interaction as fixed effects, and subject and item intercepts as random effects (BPVS standardised scores were removed for convergence reasons after confirming through a separate analysis that they had no effect on performance rather than removing items intercept which comparison between t-test results and mixed model showed to have significant impact). The control non teaching condition was used as the reference level. As expected, there was a bias for basic-level answers in the random condition of experiment 5a (est=-10.95, se=1.80, z=-6.06, p<0.0001), but no effect of teacher condition in Experiment 5a (est=0.01, se=1.48, z=0.01, p=0.995) or Experiment 6a (est=-0.14, se= 0.52, z=-0.26, p=0.791) once items and subjects were taken into account. There was also no interaction with bilingual status, and no other effects were significant.

5.4.4 Discussion and follow up

We were surprised to find no significant effect of condition in either Experiment 5a or 6a. There was a numerical trend for such an effect of condition in Experiment 5a, but a mixed model showed that this trend was not statistically robust when random effects for subjects and items were taken into account.

This failure to find an effect of condition could have occurred for several reasons. Firstly, for the action teaching task, a major difference between ours and Imai et al., (2008)'s design is that they conducted their experiment between-subject while ours was within-subject. This could have led to carry-over effects (i.e., children failing to notice the difference between conditions and consistently responding with the choice of either action or object that they had made in their first trial). Some investigation examining patterns of answering between the two lists (one of the lists started with a control trial while the other started with an action-enhanced video) indicates that this might have been the case.

Secondly, when we implemented a computer version of Xu and Tenenbaum (2007a)'s experiment, we made several changes which might have had a significant impact on answers. One major difference was that the control condition in Xu and Tenenbaum's study was a 'learner-driven' rather than random condition. In this condition, the learner was first shown an exemplar of the target subordinate category, then asked to *themselves* point to two more examples of the label used and promised a reward (sticker) to maximise conservatism in choosing an object from the same subordinate category (the 'safe' choice that would ensure reward regardless of whether the participant thought that the true extension of the label could be at the basic-level). While we believed that the suspicious coincidence effect should apply whether the two examples in the non-teacher condition were chosen by an 'ignorant' learner or provided randomly, this prediction had not been tested before. If this 'random versus teacher' effect exists, it would expected to be of smaller size than the 'learner-hoping-for-reward versus teacher' effect. Indeed, two randomly sampled subordinates should still be considered more informative with respect to the true meaning/distribution of the category than an 'ignorant' learner's choice driven by an effect entirely unrelated to the true meaning/distribution (i.e, conservatism to maximise the opportunity of a reward).

Furthermore, while we expected a basic-level bias, the percentage of subordinate answers was low even in the teacher condition. We hypothesised that this could be due to the form of the test prompt which might have biased participants towards a basic-level interpretation through the use of the word 'other'. Indeed, 'Can you touch the *other* skols?' could have been interpreted as exclusive (i.e., not the ones like the *skol* you have just seen).

Finally, another difference between ours and Xu and Tenenbaum's study was that the two subordinate exemplars in teacher condition in our study were presented one after the other rather than simultaneously, which has been shown to negatively impact the suspicious coincidence effect (Lewis & Frank, 2018; Spencer, Perone, Smith, & Samuelson, 2011).

For these reasons, and also to investigate whether the relative bilingual advantage found in experiment 1 would replicate numerically in a different and better matched sample (i.e., from a single school and of same age), we conducted follow-up experiments after making some changes to the experimental designs according to the above mentioned concerns. However, the smaller sample size and use of between-subject designs in both tasks reduced statistical power, which is why, after confirming that there were no significant differences between the results for the two experiments (i.e., no effect of task), we collapsed both ostensive teaching experiments in our analyses.

5.4.5 Experiment 5b: subordinate category teaching with learner-driven condition

The purpose of this experiment was to investigate whether the relative bilingual advantage evidenced in experiment 1 would be numerical in a better matched sample of bilingual and monolingual children.

The task was again adapted from Xu and Tenenbaum (2007a) and implemented as a computerised task but this version was constructed to be more similar to the original study. This meant that instead of having a 'random' condition where the exemplars were provided randomly, we created a learner-driven condition where the participant was given an exemplar of the subordinate category and then provided the second exemplar *themselves*. A reward incentive ensured that they would be conservative and choose a subordinate rather than basic-level exemplar while feedback ensured that the information given was the same as in the teacher condition(ie., two positive exemplars of the subordinate category).

Participants

Participants were the same 120 children as in experiments 3a and 3b (cf. Chapter 4). One bilingual and four monolingual children were excluded from analyses for not completing both tasks. The task designs were between-subject and participants were randomly assigned to critical or control condition. There were 60 children in each condition.

Stimuli

Stimuli were the same as in Experiment 5a, with the changes we mentioned. An example stimulus for the teacher condition can be found in figure 5.5, and for the learned condition in figure 5.6.

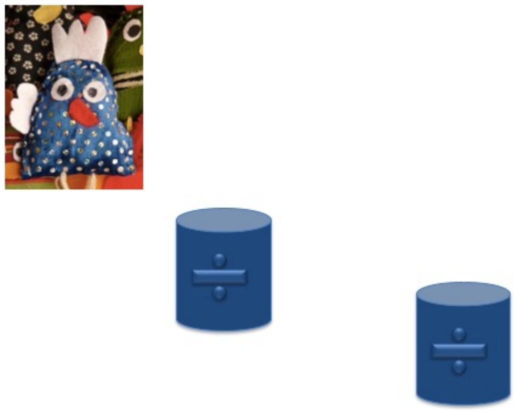


Fig. 5.5 Example stimulus experiment 5b (teacher condition), *'Mr Puppet wants to show you a dax. And look, Mr Puppet wants to show you another dax!'*

Procedure

Procedure was the same as in Experiment 5a except for three differences. First, exemplars in teacher condition were presented simultaneously rather than one after the other: in the training phase of the teacher condition, children saw an image of Mr. Puppet along with an object from the target subordinate category and heard a prompt of the type *'Mr. Puppet wants to show you a skol. Look, this is a skol!'*, another exemplar of the same subordinate category then appeared next to the first one and the children heard another prompt of type *'And, look, Mr. Puppet wants to show you another skol!'* (see Figure 5.5).

Second, the random condition was replaced by a learner-driven condition as in Xu and Tenenbaum's (2007) original study: in the training phase of the learner-driven condition they first saw a picture of the target subordinate category and heard a prompt of the type *'Oh, look what fell out of the bag, it's a liput! A liput fell out of the bag!'* and then were asked to 'pick one more' among the three subordinate categories on the picture and told that they would receive a sticker if they were right (see Figure 5.6).

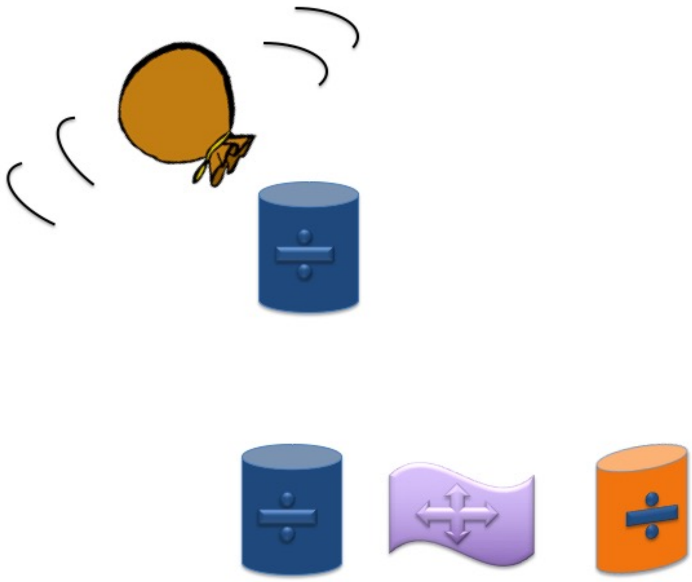


Fig. 5.6 Example stimulus experiment 5b (control condition), *'Look, this is a dax. Can you pick another dax? If you get it right you get a sticker'*

Finally, the test prompt *'Can you touch the other skols?'* was replaced with *'Can you touch the skols?*.

Results

As in Xu and Tenenbaum (2007a), a reward (sticker) had been promised in the learner condition of experiment 5b to maximise conservatism in choosing an object from the same subordinate category and the few children (n=3) who chose a different subordinate at training time were excluded from analyses. This number of non-conservative choosers (i.e., participants who did not choose exemplars from the same subordinate in the training phase of the learner-driven condition) is higher than in the original study (n=1) but similar to the percentage of adults who made this choice in Lewis and Frank (2016)'s replication of Xu and Tenenbaum's study (n=21 out of 296).

Answers in the subordinate category teaching were coded as 'basic level' if all four objects from the target category had been selected and as 'subordinate level' if only the two objects from the target subordinate category had been selected. Answers which did not fit either pattern of response (30%) were excluded from analyses. As expected, there was a bias for basic-level answers in the control (learner-driven) condition of experiment 5b (t=-2.18, df=180, p=0.03) but no bias in control (action neutral) condition of experiment 6b, where object and action answers were equally likely (t=-0.13, df=219, p=0.89).

We ran a logistic regression mixed model to investigate the reliability of these results (D. Bates et al., 2014) with items and participants as random effects and starting with all design-relevant fixed effects as random slopes, i.e., a maximal random effects model. Following a procedure used in cases of over-parameterisation (Cane et al., 2017; Horowitz et al., 2018), random effects were removed only when they led to non-convergence. Model included task, gender (shown to have an effect on pragmatics, cf. Stiller et al., 2015), bilingual status, normalised BPVS scores, condition, and their interaction as fixed effects, and subject and item intercepts as random effects.

There was a significant bias towards basic-level choices (est=1.18, se=0.23, z =5.14, p <0.001) and significant effect of condition, with subordinate choices significantly lower in learner than teacher condition (est=-1.62, se =0.38, z =-4.96, p <0.001). There were no other significant effects.

5.4.6 Experimend 6b: action word teaching between-subject

The purpose of this experiment was to investigate whether the relative bilingual advantage evidenced in experiment 1 would be numerical in a better matched sample of bilingual and monolingual children.

The task was the same as in Experiment 6a but was conducted between rather than within-subject to avoid carry-over effects.

Participants

Participants were the same 120 children as in experiments 3a, 3b (cf. Chapter 4) and 6a.

Stimuli

Stimuli were the same as in Experiment 6a.

Procedure

Procedure was the same as in Experiment 6a.

Results

We ran a logistic regression mixed model to investigate the reliability of these results (D. Bates et al., 2014) with items and participants as random effects and starting with all design-relevant fixed effects as random slopes, i.e., a maximal random effects model. Following a procedure used in cases of over-parameterisation (Cane et al., 2017; Horowitz et al., 2018), random effects were removed only when they led to non-convergence. Model included task, gender (shown to have an effect on pragmatics, cf. Stiller et al., 2015), bilingual status, normalised BPVS scores, condition, and their interaction as fixed effects, and subject and item intercepts as random effects. No effects were significant.

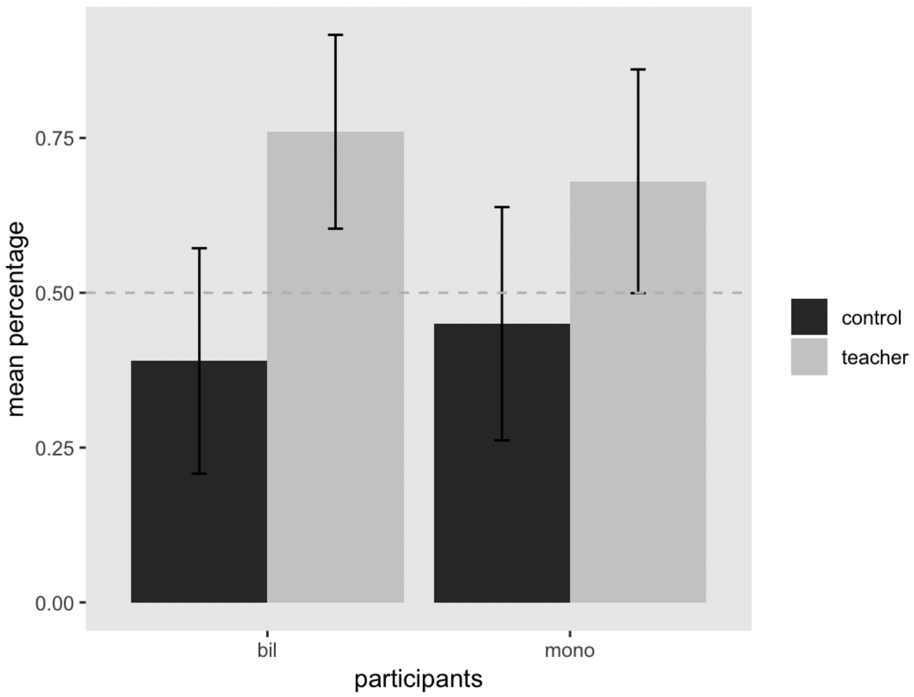


Fig. 5.7 Results experiment 5b (subordinate category teaching)

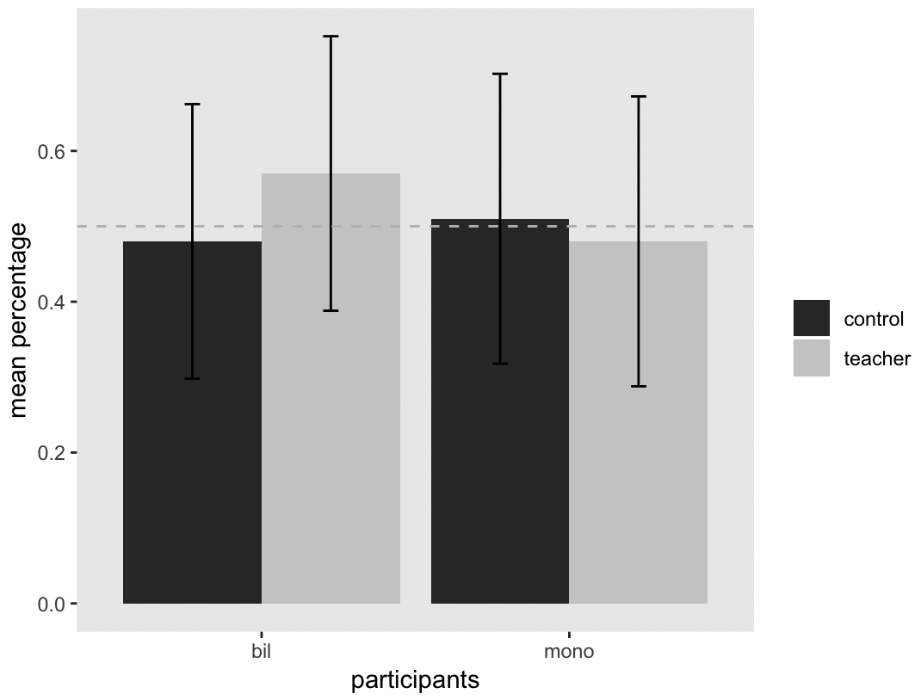


Fig. 5.8 Results experiment 6b (action word teaching)

5.4.7 Collapsed results

Results showed that target choices (subordinate category or action) were higher in teacher condition in both experiments. Additionally, bilinguals' target choices in teacher condition were higher than monolinguals' in both tasks, and bilinguals, contrary to monolinguals, were significantly above chance in the action teaching task ($t=2.02$, $df=135$, $p=0.04$) (see table 5.2). A graph showing the results of the subordinate teaching task can be seen in figure 5.7 and the results of the action teaching task in figure 5.8.

As mentioned in chapter 4, preliminary analyses revealed that the bilingual group had significantly lower English vocabulary levels ($m=65.18$, $sd=15.16$) than the monolingual group ($m=79.09$, $sd=17.01$) ($t = 14.95$, $df = 499.2$, $p<0.001$). BPVS normalised scores were then entered in the regression model as a covariable. Answers in the two tasks were collapsed after confirming that there was no significant effect of task. Bilinguals' scores in the non teaching condition ($m=0.44$, $sd=0.50$) were numerically similar to monolinguals ($m=0.48$, $sd=0.50$), however they performed higher in the teaching condition ($m=0.66$, $sd=0.47$) (monolinguals: $m=0.57$, $sd=0.49$). We ran a logistic regression mixed model to investigate the reliability of these results (D. Bates et al., 2014) with items and participants as random effects and starting with all design-relevant fixed effects as random slopes, i.e., a maximal random effects model. Following a procedure used in cases of over-parameterisation (Cane et al., 2017; Horowitz et al., 2018), random effects were removed only when they led to non-convergence. Model included task, gender (shown to have an effect on pragmatics, cf. Stiller et al., 2015), bilingual status, normalised BPVS scores, condition, and their interaction as fixed effects, and subject and item intercepts as random effects. The control non teaching condition was used as the reference level. There was a significant effect of condition ($est=1.04330$, $se=0.41$, $z= 2.52$, $p=0.01$), with target choices significantly higher in

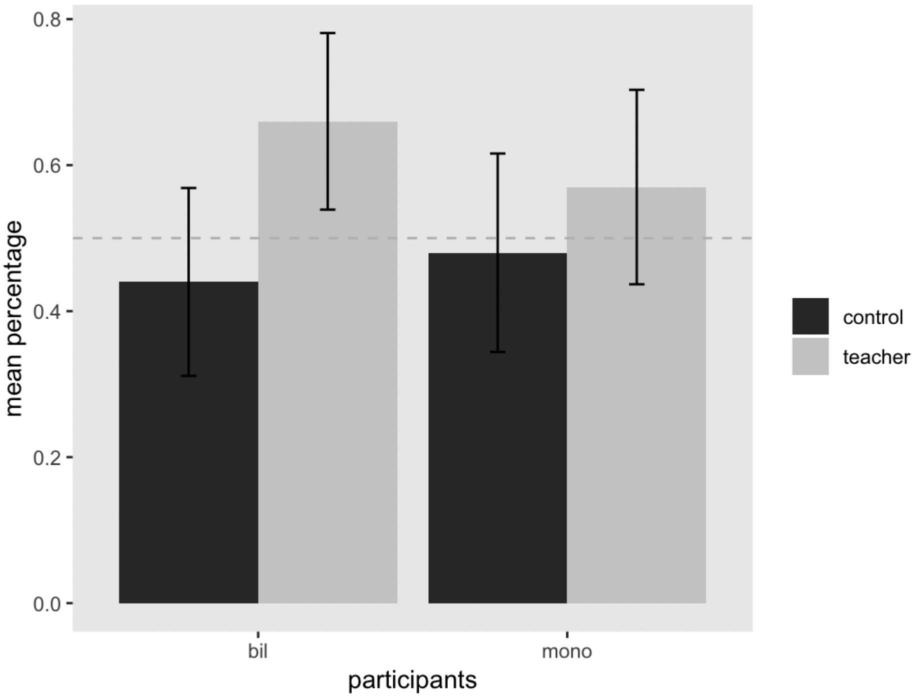


Fig. 5.9 Results ostensive teaching

teacher condition, and a significant interaction between bilingual status and condition, with monolinguals' performance significantly lower than bilinguals in the teaching condition (est=-0.74, se= 0.37, z= -1.981, p=0.04). No other effects were significant.

Table 5.2 Experiment 4: summary of results

Condition	Monolinguals		Bilinguals	
	mean	SD	mean	SD
(exp5b) subordinate category (control)	0.45	(0.49)	0.39	(0.50)
(exp5b) subordinate category (teacher condition)	0.68	(0.47)	0.76	(0.43)
(exp6b) action verb (control)	0.51	(0.50)	0.48	(0.50)
(exp6b) action verb (teacher condition)	0.48	(0.50)	0.57	(0.50)
average control	0.48	(0.50)	0.44	(0.50)
average teacher condition	0.57	(0.49)	0.66	(0.47)

5.4.8 Discussion

The goal of this study was to investigate the use of ostensive teaching for word learning by children who were monolingual or regularly exposed to another language. While children can use pragmatic inference to 'fast map' a novel word to a referent when the primary goal of the communication is not overtly didactic (as in Experiment 1, 2 and 3), humans have developed more specific and direct ways to convey referential content and allow the learner to be more confident in subsequently extending the novel label according to the teaching they have received (cf., *natural pedagogy*, Csibra & Gergely, 2009). Overall, we found that children were able to make use of the ostensive cues to extend the meaning of the novel label, as their choices of the taught (subordinate or action) category were significantly higher in teacher than in control condition. However, contrary to monolingual children, billinguals were able to reliably extend a novel label to a novel highlighted action in experiment 6b, and bilinguals' target choices in teacher condition were significantly higher than monolinguals in teacher condition, despite a significantly lower structural language ability (as measured by passive vocabulary scores). Importantly, the cues provided with the novel words were the same in teaching and control conditions in both experiments (i.e., two exemplars from a subordinate category/ a new word uttered while performing a novel action with a novel object) but a crucial difference was the ostensive context leading to these cues. Another important feature of this paradigm is that the ostensive cues given could not be used directly (as point or eye gaze could be) to select the intended referent. Instead, the child had to reason about *why* these cues had been provided, i.e., about the intentions of the speaker/teacher in order to select the correct referent for the new word.

As a consequence, even though the tasks were likely to involve ignoring a bias towards non-target answers (extension to object or basic-level), participants could only overcome this bias by reasoning about the intentions behind the context of each word teaching instance rather than by simply following a salient point or eye gaze. Going back to the three hypotheses which could potentially explain the better performance of bilingual children in socio-pragmatic word learning paradigms, our results could not have been solely due to better inhibitory abilities or to increased attention to ‘direct’ and salient social cues such as eye gaze or point, but only to a better ability to perform true pragmatic reasoning for inferring the referent of a novel word.

There are several developmental effects found in the literature both in the learning of action words and in learning hierarchical levels from ostensive teaching. In Imai et al. (2008), Chinese children were unable to extend a novel label to a novel action before 8 years old, even when it was presented as a verb with arguments, and displayed a strong bias towards object choices (3 year olds gave 33% of action answers and 6 year olds 52%, with no differences between bare word and verb with arguments), whereas Chinese adults actually displayed an action bias for bare words (73% action choices). Time of exposure to the object during the training video appeared to have a strong effect on object bias: when a short object-holding sequence was removed from the training video, 5 year olds became able to reliably extend a novel label to a novel action in a ‘verb with argument’ condition (88% action choices). This might contribute to explain why there was no object bias in our study, as the videos in our ‘action neutral’ control condition were slightly shorter, which resulted in shorter exposure to the novel object.

This effect might also explain why monolinguals’ action choices in the ‘action highlighted’ teacher condition were similar (or even slightly lower) to those in the control condition, as the teacher video was of greater length, resulting in object exposure which was on average twice as long as in the control condition, whereas exposure to target action (which shared time with the distractor action) was about the same in both conditions. Thus, if relying on pure salience (and without reasoning about which referent was more novel and thus more likely to be referred to), longer exposure to object in the 'action highlighted' video could have led to more object choices, while the highlighting of the action should lead to more action choices. If these effects were of approximately similar strength, they could potentially cancel each other and result in action choices remaining at chance level, which is what we see in monolinguals.

Why did two year olds succeed in mapping novel word to novel action in Tomasello and Akhtar's (1995) study but four to six year old monolinguals failed to extend a novel word to a novel action when given similar pragmatic cues in our study? There are several possible reasons. First, Tomasello and Akhtar's study involved an act-out paradigm with real objects which both child and experimenter manipulated and played with. This type of task and setting is likely to be more engaging and to lead to increased levels of attention. In addition, the context surrounding the naming of object and action was highlighted in a much more insistent manner, with two minutes of preparation highlighting either object or action and a total of 10 instances of the new word heard by the child for each trial. Finally, our test of extension of the novel action word was more stringent and more cognitively demanding, as it involved generalising the novel word to a new object while simultaneously ignoring the original object used in teaching the action word.

Both groups of children were able to reliably extend the novel label to a subordinate category when they had been provided with two exemplars from a same subordinate category by a knowledgeable speaker/teacher. This is due to the 'suspicious coincidence effect', cf. Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007, 2009) whereby two exemplars of the same subordinate category (e.g., two dalmatians) are considered relatively unlikely to have been randomly sampled from a basic-level distribution (e.g., dogs) and even more unlikely to have been picked as a representative sample by a knowledgeable speaker with didactic intent (i.e., who wants to teach a hearer that the new word extends to the basic-level, e.g., to all dogs). In addition, while bilinguals had a numerical tendency to extend the novel word according to the expected basic-level bias slightly more often than monolinguals, they were significantly more efficient in overcoming this bias in the teacher condition by choosing to restrict the extension of the label to the subordinate category only. Here again, bilinguals performed in a more adult-like way than monolinguals (in Xu & Tenenbaum, 2007a, children with mean age 4;0 extended the label to the subordinate level 71% of the time after being presented with three exemplars whereas adults did so 93% of the time).

Given the significant difference in English lexical knowledge between the two language groups, it is possible that the minimally verbal nature of the two tasks (particularly the action teaching task) actually helped in 'levelling the playing field' between the groups and gave them an opportunity to perform on equal grounds, allowing the bilinguals to display enhanced ability to use ostensive cues provided by a teacher to identify and extend the meaning of a novel word.

Contrary to experiments 1, 2 and 3, the two tasks discussed here involved not only forming a mapping between word and referent but *extending* the novel word during a test phase where training and test items were not simultaneously available (as they were in experiment 2). Participants thus needed to retain the mapping formed between the training and (immediately following) test phase after the screen had changed and was not displaying training items anymore. This could have been somewhat influenced by working memory capacity, which has sometimes been found to be higher in bilingually raised children (e.g, Morales, Calvo, & Bialystok, 2013)[1]. Because we do not have working memory measures for this sample of children, we cannot rule out this possibility. However, as we mentioned in the introduction, there is a debate about the so-called

[1]Making training items available at test time in Lewis and Frank (2016, Experiment 2) did not appear to have a significant effect on results, but this might have been due to participants being adults with better short term memories.

'executive control' advantage in bilinguals, and this advantage in working memory (as for other executive control skills) often does not replicate (e.g., Engel de Abreu, 2011).

In conclusion, while both adults and children display prior biases for the extension of a novel word, such as basic-level or object bias, these are generally stronger in children, possibly because they are still at a stage of vocabulary acquisition where these heuristics are particularly needed and helpful for constraining the hypothesis space of referents. However, compared to monolinguals, bilinguals appear to apply these biases more flexibly and to more readily agree to bypass them when given pragmatic cues to that effect by a knowledgeable teacher.

Chapter 6

Conclusions

Parents of young organic life forms should be warned that towels can be harmful if swallowed in large quantities.

— Douglas Adams

6.1 Implications for theories of bilingual language acquisition

The purpose of this thesis was to investigate the ability of monolingual and bilingually exposed children to perform pragmatic inference in the context of word learning. The reasons for undertaking such investigations were twofold. First, the thesis sought to examine the nature of a previously found advantage in socio-pragmatic word learning displayed by bilingual children (Yow, 2015; Yow & Markman, 2011a, 2011b, 2015, 2016). Second, it sought, through examining the formal aspects of this advantage, to help solve a research puzzle: the ability of bilingually exposed children to acquire vocabulary quickly and efficiently, and the precise way they achieve this.

In Chapter 2 we introduced the research questions, the main types of word learning theories (pragmatic and associative), the specific characteristics of a bilingual environment (more varied input and higher risks of miscommunication) and the puzzle of the bilingual lexicon (which is not half the size of a monolingual's despite less reliance on mutual exclusivity), as well as previous evidence of a bilingual advantage in socio-pragmatic word learning and in pragmatic skills more generally. We then described the aims of the thesis, namely to find out whether this previously found advantage in word learning involved a better ability to perform true pragmatic reasoning, i.e., reasoning about speaker's intentions, as opposed to simply better inhibitory control or better attention to social cues such as prosody, eye gaze or pointing.

In Chapter 3, we conducted a first experiment (Experiment 1) whose goal was, as we said before, to investigate the ability of a group of monolingual and bilingually exposed children to use pragmatic inference to perform fast mapping between a novel word and referent when 1) success could not be achieved by ignoring or inhibiting an irrelevant cue 2) success could not be achieved by simply making direct use of a social cue (such as following eye gaze or point). This first experiment involved a display of four characters, two of which were modified, and the need to use the presence of a stressed adjective to infer the referent of a novel word. We found that, while the bilingually exposed group had a worse performance on tasks involving purely lexical knowledge (as verified in the control condition and in their vocabulary scores), they performed significantly better than monolinguals in the pragmatic condition task when taking this baseline into account. We concluded that the bilinguals displayed a relative advantage in using pragmatic inference for fast mapping even when there was no irrelevant cue to be inhibited (and thus that results could not have been driven purely by inhibitory control ability).

This conclusion was further strengthened by the contrast between the results of the first and the results of the second experiment In this experiment (Experiment 2) we examined the ability of the same group of monolingual and bilingually exposed children to perform fast mapping using, again, a prosodic cue (emotional affect). However, a crucial difference between Experiment 1 and Experiment 2 is that emotional affect, unlike stress, is a social cue which, once it has been learned, can be used directly in an associative fashion without the need to reason about speaker's intentions. Accordingly, we found no difference in performance between the monolinguals and bilingually exposed, and performance appeared to rely on a different set of abilities/predictor variables (e.g., no effects of gender or vocabulary scores which had been found in Experiment 1). This is in line with previous research showing that, unlike it has been claimed before (cf., Yow & Markman, 2011a), there is no bilingual 'advantage' for recognising emotional affect but rather a bilingual 'bias' (Champoux-Larsson & Dylman, 2018) whereby bilinguals are biased towards an answer which is consistent with affect whether it is relevant to the task or not. From the results of Experiment 1 and 2 we concluded that the previously found differences in pragmatic word learning between monolinguals and bilinguals are not solely due to better inhibitory control or a better ability to use social cues in a direct, associative way, but to a better ability to reason about speaker's communicative intentions.

In Experiment 1 we found that children, unlike adults, were not able to perform the fast mapping task in a condition where the adjectival modifier did not also feature contrastive stress (thus providing two cues for performing the pragmatic inference). In Chapter 4, we describe another experiment (Experiment 3) which allowed us to investigate whether this difference between monolingual and bilingually exposed groups would appear more clearly in a paradigm where only one pragmatic cue was available, but this cue was based on assuming the speaker not to be *underinformative* rather than assuming them not to be *overinformative*, as in Experiment 1, which arguably might result in the inference being easier to perform. In this task, participants were asked to infer and extend the meaning of a novel word based on frequency alone. A character with two objects was described using a novel word and participants had to infer that the novel word referred to the least frequent of the two objects (i.e., the one that was not featured on the other training character). They then had to extend the word by choosing between two new characters which each had one of the objects. Since children, unlike adults, had failed at deriving a pragmatic inference from a single cue using *overinformativeness* in Experiment 1, Experiment 3 examined performance of a pragmatic inference from a single cue based on *underinformativeness*.

Given that children of the same age had been shown to succeed in a similar task before, we were surprised to find that children, unlike adults, did not perform above chance in our task and were biased towards choosing the non-target object. We concluded that this was due to our paradigm introducing an unexpected salient alternative and conducted further follow up experiments to investigate this possibility, which we discuss in the next section. However, we noted that, while neither the monolingual nor the bilingually exposed group could perform the target pragmatic inference, bilinguals were significantly less biased than monolinguals towards the non-target object. This could have been the result of a lesser tendency to answer according to a visually salient, non-pragmatic cue, or to a better ability to pay attention to a diversity of available cues and/or particularly the presence of socio-pragmatic cues (while perhaps not being yet able to perform full pragmatic reasoning with the support of only one cue). This would be in line with other works (Colunga et al., 2012; Yow et al., 2017; Yow & Markman, 2015) which tend to indicate that bilinguals are better able to follow socio-pragmatic over visually salient cues and to integrate multiple referential cues. This also aligns with recent work on pragmatic development which suggests that inferences are performed in a constraints-based manner, with performance, especially in children (who struggle to derive them with one cue only, see next section), moderated by number, strength, and reliability of cues (e.g., Degen & Tanenhaus, 2015).

In view of the results of Experiment 4, we wished to examine 1) whether children could perform a pragmatic inference on the basis of a single cue in an ostensive teaching setting where the didactic/communicative intention was clearer 2) if bilingual children could, in such a setting, use the pragmatic reasoning ability demonstrated in Experiment 1 to ignore a default response and correctly extend the meaning of a novel word, 3) whether a better matched (in age and socio-economic background) sample of bilingual and monolingual children would this time display not just a relative but a numerical advantage and 4) whether the relative bilingual advantage found in Experiment 1 would apply not just to fast mapping but also to extending a novel word. To this end, in Chapter 5 we present the results of two further experiments (Experiments 5 and 6) conducted with a different sample of monolingual and bilingual children. In the latter experiments, participants were asked, after being given a pragmatic cue by an ostensive teacher, to extend the meaning of a novel word to the subordinate rather than basic level (Experiment 5) or to a novel action rather than a novel object (Experiment 6). In previous work by Yow and colleagues, bilinguals were better able to ignore a default or salient response by focusing on a salient social cue (such as eye gaze or point). In our Experiment 1, we showed that bilinguals displayed a relative advantage in making use of a pragmatic cue (stressed adjective) for fast mapping when there was no need to inhibit an irrelevant cue. Here we wanted to see whether bilinguals could *use* this pragmatic reasoning ability in order to inhibit a default response and extend the meaning of the novel word to a new instance of the referent. Crucially, unlike in previous work by Yow and colleagues, success (or inhibition of the salient response) could not be achieved by simply using another (salient) social cue in a direct way, but participants had to perform true pragmatic reasoning about speaker's intentions in order to achieve it. In line with our predictions we found that 1) children were able to use a single pragmatic cue for word learning and extension in a supportive, ostensive teaching context, 2) bilinguals were significantly better than monolinguals in using ostensive pragmatic cues to inhibit a default response in word learning, 3) in a better matched sample, this bilingual advantage became a numerical one and 4) this advantage applied in a paradigm which required not only fast mapping but also extending the meaning of the novel word.

What wider implications can we draw from these results? It might be useful to go back and consider the reasons we study bilingualism and bilingual individuals. First, since a bilingual environment is in many parts of the world the rule rather than the exception, it is of crucial importance to study both the short and long term impacts of multilingual input on language acquisition and socio-cognitive development more generally. This is done in the hope to verify that there are no long-term detrimental effects on language and other abilities, or that the overall balance taking into account socio-cultural and emotional benefits is a positive one. Failing this, such studies give parents and educators the opportunity for making a conscious and informed choice.

Second, bilingual input should act as a compulsory and stringent test for any model of language acquisition or language processing. In general, multilingual input provides an alternative environment for testing theories of language acquisition, for example when theories make different predictions regarding syntactic transfer effects (Paradis & Genesee, 1996). In our case, as we mentioned in Chapter 2, even hybrid computational models of word learning (i.e., using both pragmatic and cross-situational information) fail to mimick the performances witnessed in children when faced with multilingual input (Zinszer et al., 2018)). Thus, bilingual studies constitute an efficient and practical way of challenging assumptions made by a number of linguistic theories and models.

Third, bilingualism offers a privileged ground for examining how experience and the constraints of a specific environment contribute to shaping the development of cognitive and social skills. More specifically, bilingual studies can shed light upon which skills are indeed impacted by certain environmental variables, and which are not. Humans are famously good at adaptation and flexible learning (Fazey et al., 2007). However, the extent and nature (in terms of domain-generality) of each postulated adaptation needs to be thoroughly investigated and replicated empirically before any definite conclusions can be reached. For example, while it might be intuitively appealing to endow the visually impaired with better hearing, this does not appear to apply to general auditory abilities (e.g., differential auditory sensitivity and acoustic reflex thresholds) (Starlinger & Niemeyer, 1981). However, blindness does seem to have a significant positive impact on echolocation skills (Schenkman & Nilsson, 2010). Here we see a contrast between changes in a very general ability, versus a specific skill which might result from the existence of an imperative and primary need (i.e., to move safely from one place to another). Once again, as organisms who depend on efficiently managing a limited amount of energy, we appear to use and develop resources on an 'as-needed' basis; no more and no less than is required.

Language acquisition and understanding arguably constitute some if not the most important enablers of social interaction and knowledge transmission. As such, their facilitation is likely to be prioritised, and adaptive compensatory strategies to be developed in environments which are linguistically 'hostile' or challenging. One such example are noisy environments, giving rise to the 'cocktail party effect', whereby hearers strain to process speech in the presence of interfering auditory background (Cherry, 1953; Yost, 1997). Conversational partners in such situations tend to display different pragmatic strategies in the form of changes in discourse structure and patterns of interaction. While regular pragmatic behaviour in conversation involves the cooperative development of themes, general coherence, cohesion and efficient turn-taking, the pragmatics of noisy environments present a number of patterns which bear striking resemblance to the strategies used by the deaf and hard of hearing (D/HH) in interactions with hearing peers: avoidance of talk, requests for clarification, for repair and repetition, for louder or slower speech, body repositioning, less discussion of shared topics, shorter turns or monologues and 'artificial' acknowledgments (i.e., pretense of comprehension) (McKellin, Shahin, Hodgson, Jamieson, & Pichora-Fuller, 2007). Visual cues and lipreading can also help support speech comprehension in environments with high levels of background noise (Sumby & Pollack, 1954).

In the face of this evidence, it would seem reasonable to assume that people who are hearing impaired might have enhanced pragmatic and visual speech processing abilities. However, while D/HH do demonstrate reliability enhanced lipreading skills (AuerJr & Bernstein, 2007), face recognition skills (De Heering, Rossion, & Maurer, 2012) and even general visual processing (i.e., texture segmentation and visual search) under certain conditions (Rettenbach, Diller, & Sireteanu, 1999), this does not appear to be true for pragmatic skills. Indeed, D/HH children were found to perform no better than hearing children on detecting violations of conversational maxims (Surian, Tedoldi, & Siegal, 2010) and to be generally at equal levels of performance as children with ASD (Autism Spectrum Disorders) on theory of mind and false belief tasks, results which were believed to follow from a significantly lower amount of conversational experience (Peterson & Siegal, 1995). Despite this, D/HH do display a wider range of pragmatic skills than their hearing peers, asking more questions, taking longer turns and initiating more topics (Paatsch & Toe, 2013). By contrast, ASD children are more likely than neurotypicals to respond inappropriately to requests for clarification (Volden & Phillips, 2010) and have a reduced ability to perceive speech in noisy environments (Alcántara, Weisblatt, Moore, & Bolton, 2004)

What about bilingualism? It might be argued that bilinguals, like the hearing impaired, evolve in an environment which is linguistically and communicatively challenging. However, unlike D/HH children, bilingual children do not experience a significantly reduced amount of conversational opportunities. Because of this, it might be expected that they would be able to display compensatory abilities in related domains, some of which would naturally result from the higher variety they face in their input. For example, bilingual children show less perceptual narrowing (Byers-Heinlein & Fennell, 2014) and are more accurate at face recognition (Kandel et al., 2016). Bilingual 8-month-old infants, unlike monolinguals, are able to distinguish between two silent videos of unknown languages (Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012) including bimodal bilinguals (ASL/English) (Palmer, Fais, Golinkoff, & Werker, 2012) In addition, bilingual 12-month-olds, unlike monolinguals, continue to monitor mouth movements for both native and non-native speech (Pons, Bosch, & Lewkowicz, 2015).

These differences between bilinguals and monolinguals are consistent with an effect of type of input (perceptual narrowing) and an 'as needed' basis, since bilinguals are likely to have a greater need for face recognition (as a language cue), for distinguishing between languages and for monitoring speakers' mouths (at least in early stages of language acquisition) to support weaker phoneme discrimination due to lesser exposure in each language (Fennell, Byers-Heinlein, & Werker, 2007). In contrast, bilinguals do not perform better at processing audio-visual speech and benefit less than monolinguals from the addition of visual cues compared to an audio-only condition (Kandel et al., 2016). This is consistent with other works showing a general delay in word recognition for bilinguals (Gollan, Salmon, Montoya, & Galasko, 2011; Shook, Goldrick, Engstler, & Marian, 2015) , an effect which results from both lack of exposure and a greater number of potential candidates and would be expected to apply to both audio and visual mediums. Indeed, contrary to an intuition of 'inverse effectiveness' whereby visual cues would be most effective as SNR (Signal-to-Noise) ratios decrease, these cues are actually most effective at intermediate levels of sound quality (i.e., they are useless at levels that are too low and not needed at very high levels) (Kandel et al., 2016). This means that humans, which appear to behave in a near Bayes-optimal manner in combining cues for a variety of multisensory tasks such as audiovisual speech (Kandel et al., 2016), visual-tactile integration (Rach & Diederich, 2006) and texture/motion depth assessment (Jacobs, 1999) benefit most from several cues of intermediate strength. And, as Marian (2009) points out, SNR ratios are likely to be lower in bilinguals because of their decreased ability to perceive sound contrasts (Bent & Bradlow, 2003).

The worse performance displayed by bilinguals in audiovisual speech interpretation would tend to suggest that their environment does not foster a general advantage in integration of linguistic and paralinguistic cues, at least not of cues for which monolinguals have a net advantage resulting from greater exposure. However, as we said previously, there is no principled reason to believe that this should be the case, as the effect of experience might not extend to domain-general abilities and, unlike visual language discrimination, there is no particular bilingual need for better audiovisual processing.

Going back to the three hypotheses we put forward for explaining the previously found bilingual advantage in socio-pragmatic word learning, we do not further examine the first (i.e., better inhibitory control or better selective attention), precisely for the same reasons mentioned above, namely that the exertion of a specific competence (inhibiting and switching between languages) may not lead to broader changes in all aspects of cognition (Warmington, Kandru-Pothineni, & Hitch, 2018) and because in such a heterogeneous population the amount of exertion of this competence may vary widely depending on the type of bilingual environment (home/outside home, one person/one language, etc.)(Green & Abutalebi, 2013). It will suffice to say that the results of our Experiment 1 suggest that this advantage is not purely due to better inhibition of an irrelevant cue.

Turning to the second hypothesis (i.e, increased attention to social cues), if we assume that this skill does not involve any theory of mind or intention reading, we are forced to classify it as a special case of the inhibition/attention hypothesis. In this view, rather than (or in addition to) enhanced domain-general executive abilities, bilinguals would also have developed a specific sensitivity to social (speaker-provided) cues such as eye gaze or pointing. This would be the consequence of experiencing higher risks of communication failures (Wermelinger et al., 2017) and a higher level of uncertainty in the signal provided by the speaker, for example frequent code-switching (Yow & Markman, 2016). As a result, these cues would prove particularly salient for bilinguals, who would be biased to pay particular attention to them compared to other types of cues. As we said, while we do not rule out this possibility, the results of our experiments show that this previously found advantage cannot be due only to increased sensitivity to social cues and the ability to use them directly to find the referent of a novel word.

Coming to our third and final hypothesis (i.e., true pragmatic reasoning which involves intention reading), two questions arise about the research undertaken in the current thesis. First, if the intention-reading is the component distinguishing between second and third hypotheses, why not directly study perspective-taking and theory of mind as the means used by bilingual children to achieve efficient word learning? Second, why should we postulate that the bilingual environment would foster pragmatic skills for word learning over and above the enhanced sensitivity to social cues described in the second hypothesis?

To answer the first question, we draw attention to work described in the introduction which shows that the ability to perform intention-reading (e.g., in ASD individuals) in a task which explicitly examines this competence does not necessarily entail the ability to use this skill in a naturalistic context, for example to cancel a pragmatic implicature (Hochstein et al., 2018). We would further extend this idea to claim that differences which may not be consistently apparent in general pragmatic inference paradigms might become clearer in the context of word learning, as per the previously mentioned 'as needed' principle.

To answer the second question, we turn to the word learning model described in Frank (2014) and Frank, Goodman and Tenenbaum (2009). This model combines both inferred meanings (pragmatic reasoning) and cross-situational associations (sometimes achieved using social cues such as eye gaze and pointing) in a parallel manner using Bayes' rule, and achieves far better performance than models which use only one source of information. Thus, it is likely that the higher risks of miscommunication and uncertainty in input would drive an increased use of intention reasoning as an added cue for word learning. As we have seen, inference seems to proceed through near-Bayes-optimal cue integration (cf., Kandel et al., 2016; Rach & Diederich, 2006; Jacobs, 1999). This could explain why speakers and hearers are often found not to behave in an optimally Gricean way (Engelhardt et al., 2006; Rubio-Fernandez, 2016; Horton & Keysar, 1996; Lane & Ferreira, 2008), as they might only prove optimal when integration of all available contextual cues is taken into account. We come back to this idea in the next section. If we define pragmatic competence as precisely the ability to perform cue integration of both intentional and other contextual sources, and given the evidence from ours and other works that bilingualism increases flexible use and balanced weighting of these cues, we might hypothesise that bilinguals benefit from enhancements in both aspects of pragmatic skill, i.e., intention reasoning and cue integration involving intentions. Indeed, pragmatics is a domain which is likely to be less strictly tied to exposure. Further work will have to examine this claim. It would appear that a bilingual environment makes children more cautious and prompts them to weigh all cues available rather than 'jump to conclusions'. By contrast, it was hypothesised that the struggles of ASD children (who are known to experience pragmatic difficulties) in noisy speech perception might be due to less effective cue integration from discontinuous sources (Alcántara et al., 2014).

We conclude that this thesis provides evidence for differences in the processing of pragmatic cues by bilingual and monolingual children which are not due solely to better inhibitory skills or to a general sensitivity to social cues such as prosody, eye gaze and pointing, but to performing true pragmatic inference by reasoning about communicative intentions in the context of word learning. We believe that these differences are the result of adapting to the challenging aspects of a bilingual learning environment, such as higher risks of miscommunication and the need to efficiently and quickly acquire words from complex and variable input. We finish by adding that making this distinction between using social cues and reasoning about intentions might help provide insights about separate developmental timelines for exerting different types of pragmatic competence, with early abilities demonstrated by the bilingually exposed, particularly in acquisition contexts.

6.2 Implications for theories of pragmatic inference

In the introductory chapter we began with the question 'what is pragmatics?'. If we adopt a cognitive and epiphenomenal perspective (Perkins, 1998):

> The consequence of interactions between a set of linguistic and non-linguistic cognitive subsystems which determine the crucial balance between how much information is encoded

linguistically and how much is left unsaid on the grounds that it is recoverable from the linguistic and non-linguistic context of the utterance.

Pragmatics is thus very much the study of pragmatic *inference*, rather than signal processing. In a communicative context where interactions take place between intelligent and purpose-driven beings, we can use assumptions of informativeness to retrieve meaning: efficiency or specificity (i.e., at equal cost, assume the most specific meaning) and relevance or coherence (i.e., assume the most related or closely associated meaning). In this thesis we focused on specificity inferences as inferences which require a lesser amount of conceptual and world knowledge.

In view of this, our first and second experiments (Experiment 1 and 2) allowed us to explore whether word learning inferences previously described as 'socio-pragmatic' could involve two distinct sets of skills underpinned by different processes, predictor variables and displaying different types of individual patterns. We showed how a similar prosodic cue would lead to differences in performance between a group of monolinguals and bilingually exposed children in one fast mapping paradigm, but not in the other, and would yield a different set of predictor variables (namely, vocabulary, gender and language status in Experiment 1, but not in Experiment 2). Indeed, in Experiment 1, the pragmatic inference could be calculated and predicted quantitatively using *Size Principle* (Tenenbaum & Griffiths, 2001) and reasoning about what the speaker would use for each potential referent, i.e., at equal cost, choosing potential referents according to the specificity of the expression used (*'the wet gorp'* being more specific in the case of several aliens of the same type). Thus, the cost of using a modifier and prosodic stress was assumed to be contextually justified, and predictions were made according to this assumption. By contrast, no such calculations on the basis of cost incurred by the speaker could be made in Experiment 2, where emotional affect (positive or negative tone) represented a cost which was constant, and thus had to be recognised and associated with an object's valence (positively or negatively modified). Interestingly, these two sets of skills (pragmatic and associative) are also the ones which underlie the competing theories of word learning outlined in the introduction. As we said in the previous section, we expect, in accordance with a hybrid model such as Frank, Goodman and Tenenbaum's (2009), that both skills would be involved in cases of 'socio-pragmatic' word learning, i.e., learning words by using 'direct' social cues such as eye gaze or pointing. We also expect associative abilities, as a lower-level competence, to be acquired and used earlier, which would explain some of the discrepancies between young children's successful performance in socio-pragmatic word learning as opposed to false-belief tasks (Wellman & Bartsch, 1988; Wimmer & Perner, 1983) and pragmatic inferences such as scalar implicatures (Noveck, 2001; Guasti et al., 2005; Huang & Snedeker, 2009). In our case, children in Experiment 1, unlike adults, appeared unable to perform a contrastive inference by reasoning about the communicative intentions of the speaker on the basis of the adjectival modifier alone (i.e., *'the wet gorp'* vs. *'the WET gorp'*).

As stated in the introduction, we do not wish to say that even very young children are never able to attribute mental states to others. Indeed, recent work using spontaneous response such as violation-of-expectation (VOE) or anticipatory-looking (AL) have shown that infants responded according to false belief attribution (Song & Baillargeon, 2008; Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007). Rather, what ours and other works seem to suggest is that developmental differences arise from differences in the ability to make spontaneous use of pragmatic cues which may not have a high level of salience, especially when other, more salient cues are available. This explains why paradigms which enhance attention to these cues show children succeeding in false belief tasks (Rubio-Fernandez & Geurts, 2013; 2016), scalar implicatures (e.g., Pouscoulous et al., 2009; Katsos & Bishop, 2011) and other types of pragmatic inference where success in children is only achieved when a number of converging cues is provided (Kurumada & Clark, 2017; Horowitz & Frank, 2014). Similarly, as we said before, individuals with ASD who were able to reason about knowledge states in one task proved unable to use this competence to spontaneously to cancel an implicature (Hochstein et al., 2018).

This brings us to our third experiment (Experiment 3). This experiment was meant to be a slightly modified version of a paradigm in which children had previously been found to succeed (Frank & Goodman, 2012). However, to our surprise, we found that children in our experiment failed at deriving the expected pragmatic inference based on referent frequency and displayed a strong bias towards the non-target, distractor referent. Following other works (Horowitz, Schneider & Frank, 2018) and examining paradigms leading to children's successes (Papafragou & Tantalou, 2004) or failures (Noveck, 2001), we suggested that the availability of a salient cue (such as a match between question and answer, or between linguistic description and visual stimuli) disrupted pragmatic reasoning. The reason, we hypothesise, is that processing such matches constitutes a more automatic and default process than reasoning about communicative intentions, especially for children who might have similar lower-level associative abilities to adults, but less experience in using intentions to infer meanings. In addition to this, some types of inference such as reasoning by exclusion might be easier to perform and similarly provide a more easily available answer. Indeed, given the results in our version of this paradigm, we would like to suggest that children's success in Frank & Goodman's (2014) study might precisely have been due to the performance of such an exclusion inference which was not warranted in the new paradigm. This conclusion is strengthened by results such as those of Sullivan and colleagues (2017) who show that children, unlike adults, do not display the asymmetry which would be expected if they performed true scalar implicatures rather than simple exclusion inferences. While adults understand that the same mismatch between question and answer should lead to rewarding the character in one case (e.g., *'Did he color some of the stars/he colored all'*) but not in the other (e.g., *'Did he color all of the stars/he colored some'*), children do not appear to make this distinction and withdraw prizes at the same rates in both cases. Crucially, they do display the asymmetrical ratings in a case where the result is displayed in a visual way rather than described linguistically, which suggests that choices are indeed driven by surface mismatch in linguistic form. This bias towards simple associative or matching processes is reported in other works where children's rates of choices for different types of quantifiers (*some of the horses/all of the horses/none of the horses*) are significantly closer together than in adults, suggesting choices based on NP matching alone (i.e., the picture with horses only) (Horowitz, Schneider & Frank, 2018). Similarly, children appear to succeed in scalar implicature paradigms involving matching pairs and fail when faced with mismatches (cf., Noveck, 2001; Papafragou & Tantalou, 2004; Syrett et al., 2017a, 2017b). In terms of reasoning profile, children might thus behave in a more 'all-or-nothing', exclusive manner ('if P, not Q', i.e., if one answer is easily available, other cues are not considered) when performing inferences while adults (and possibly bilinguals) are more Bayesian in their probabilistic weighing of available cues. This is, again, consistent with a constraint-based model of inference whereby interpretation derives from the simultaneous integration of multiple constraints originating from different sources (cf., Hanna et al., 2003; Degen et al., 2015; Heller, Parisien & Stevenson, 2016).

This being said, our last set of experiments (Experiment 5 and 6) show that children can succeed at performing pragmatic inferences for word learning using a single cue in specific types of settings. Following Csibra and Gergely's (2009) *natural pedagogy* theory, we would like to suggest that some ostensive environments and signals prepare listeners for receiving referential or generic knowledge and allow children to allocate their attention more efficiently and make better use of a pragmatic cue. However, while both groups of children succeeded in using ostensive cues for subordinate category learning in Experiment 5, only the bilingual group succeeded in using such a cue to extend a novel word to a novel action in Experiment 6. The results of these experiments further strengthen our conclusion that the ability to make use of 'social cues' for word learning has to be distinguished as an ability which is acquired earlier and does not necessarily entail the ability to perform spontaneous pragmatic reasoning. Indeed, while very young children do appear able to learn words in 'non-ostensive' paradigms (e.g., Tomasello, Strosberg & Akhtar, 1996), it is an open question which amount of pragmatic reasoning is involved in performing word learning inferences in such paradigms (i.e., there may not be any reasoning about informativeness *per se* but mechanisms which are more akin to associative processes).

The topic of the thesis allowed us to explore several outstanding questions regarding pragmatic inference. We briefly touched upon the question of informativeness as understood through specificity ('narrowness') or through relevance ('relation') of a meaning hypothesis. We discussed specificity more particularly and ranking referents according to the *Size Principle* as well as the role of alternatives. We also examined a developmental interplay between defaults and reasoning strategies. Finally, we drew a line between the use of social cues and pragmatic inference for word learning as related but separate skills with potentially separate acquisition processes.

With a broader perspective in mind, we believe our results to bear interestingly on questions related to the impact of social experience on humans as opposed to other species. Indeed, it is controversial that other species have the ability to represent the states of mind of other intelligent entities and the use of social cues in other animals, such as point following in dogs, appears to pertain to the domain of simple associative learning (Elgier, Jakovcevic, Mustaca, & Bentosela, 2012). However, our work and those of others investigating the influence of bilingual exposure on pragmatic inference contribute to show that social experience has an effect which in humans goes beyond low-level skills such as attention and associative learning, to affect the development of mind reading and social reasoning.

In the next sections, we discuss our results from a quantitative perspective, to assess the accuracy of predictions from a Bayesian framework in modelling the type of pragmatic inference and examine each of our experiments in turn, offering potential avenues for explaining discrepancies when they arise.

6.3 Quantitative discussion

6.3.1 Experiment 1a

In our first experiment, children had to infer the target referent of a novel word from the use of a stressed or non stressed modifier (e.g., *'Touch the WET/wet gorp'*). The participants were given a display with four novel 'aliens', of which two had the property mentioned by the modifier (e.g., two dry aliens and two wet aliens). Thus a purely literal listener should have been at chance in choosing between the two aliens which semantically satisfied the modifier's requirement (i.e., between the two wet aliens), that is, each of the relevant aliens should have been chosen about 50% of the time.[1].

However, a pragmatic listener might have noticed that one of the semantically relevant wet aliens had a dry counterpart (i.e., there was only one alien of type B, which was wet, but there were both a wet and a dry alien of type A). This pragmatic listener should then have inferred that the wet alien which was not the only one of

[1]This assumes that both aliens were equally salient and thus had an equal prior probability of being chosen. We controlled for a potential bias by having two lists and switching which pair of aliens was being referred to (i.e., in one list it was the 'dry' pair and in the other the 'wet' pair in the picture stimulus, each pair involving different distractors)

its kind in the display was more likely to be referred to using the modifier than the wet alien which was the only one of its kind. The reason for this is that the use of the modifier (i.e., saying *'the wet alien'*) for the single alien is not required by the context to for informativeness, just as using the expression *'the tall glass'* when there is only one glass would be considered overinformative, as modifiers usually imply a contrast between members of the same class in referential contexts (cf., Sedivy et al., 1999). In other words, there is a cost incurred by uttering the additional word which does not increase informativeness. While speakers are not expected to always be optimally informative in their use of modifiers, especially when it involves salient features such as colours (cf., Rubio-Fernández, 2016), listeners can generally assume a reasonable level of efficiency and base their referential inferences on this assumption (Davies & Katsos, 2009, 2013). For example, in Gelman and Markman (1985), four year old children presented with a display including several members of a given category (e.g., three crayons) and one member of a different category (e.g., one record) along with statements of the type *'Show me the broken one'* were more likely (73%) to pick the non-single item in a category (e.g., the broken crayon rather than the single broken record). Three year olds, however, were overall at chance (51%). Interestingly, the three year olds performed above chance for some adjectives, which were also the ones on which the four year olds performed best (*little, broken, clean*, but no colour adjectives).

According to the quantitative predictions from a Bayesian framework such as the Rational Speech Act (Frank and Goodman, 2012), this inference can be modelled using Bayes' rule as the following

$$P(m|a) = \frac{P(m)P(a|m)}{\sum_{m' \in M} P(m')P(a|m')} \tag{6.1}$$

Where m is the meaning hypothesis and a the adjectival modifier used, $P(m)$ constitutes the prior probability to refer to a specific referent or meaning (crayon or record) and $P(a|m)$ is the likelihood of a specific referent or meaning (crayon or record) given the use of an adjectival modifier. In this framework, each meaning hypothesis should be chosen with a frequency proportional to its informativeness, or in this case its specificity. This means for each meaning we have a likelihood inversely proportional to the extension of the modified referent in context, i.e., the ratio between the extension of the modified referent and the full meaning extension.

$$P(a|m) = \frac{|a|}{|m|}^{-1} \tag{6.2}$$

Where$|a|$ is the extension of the modified objects (the number of objects of a certain type which have the mentioned property) and $|m|$ the extension of all objects for this meaning (i.e., all objects of a certain type). If we consider a uniform prior (i.e., that no referent or meaning is a priori more likely) we have

$$P(m|a) = \frac{\frac{|a|}{|m|}^{-1}}{\sum_{m' \in M} \frac{|a|}{|m'|}^{-1}} \tag{6.3}$$

Thus in Gelman and Markman (1985) we would have

$$P(crayon|broken) = \frac{\frac{|broken|}{|crayons|}^{-1}}{\frac{|broken|}{|crayons|}^{-1} + \frac{|broken|}{|records|}^{-1}} = \frac{\frac{3}{1}}{\frac{3}{1} + \frac{1}{1}} = \frac{3}{4} \tag{6.4}$$

Because the modifier in Gelman and Markman is three times as informative for the target than for the distractor referent (i.e., the property of ‘brokenness’ is three times more specific for the broken crayon among the three crayons than it is for the single broken record), the broken crayon should be chosen three times more often than the broken record (75% vs. 25%). As we can see, this prediction is very close to the numbers found in the 4 year olds (73%). In comparison, in our study, as there were only two members in the category of the target referent (i.e., only two aliens of type A), targets should be chosen about twice as often as distractors (67% vs. 33%)

$$P(alienA|wet) = \frac{\frac{|wet|}{|aliensA|}^{-1}}{\frac{|wet|}{|aliensA|}^{-1} + \frac{|wet|}{|aliensB|}^{-1}} = \frac{\frac{2}{1}}{\frac{2}{1} + \frac{1}{1}} = \frac{2}{3} \tag{6.5}$$

Again, this is very close to the numbers we obtained in adult participants in the stressed condition (66%) when removing incorrect choices (i.e., dry alien choices when participants were asked to pick a wet alien). It is worth noting that we are only able to compare Gelman and Markman's results to the results in our stressed condition, as the authors indicated that primary stress was also placed on the adjective in their study. From the results in both adult and children participants, it seems like the RSA's rationality parameter α which modulates the strength of the pragmatic inference should be adapted in non-stressed cases, i.e., if listeners are expected to be maximally rational (α=1) when hearing stressed modifiers, α should be set closer to 0.88 when modifiers are not stressed. It is interesting that this ratio appears to be almost exactly the same in adults and children.

However, children's performance in our stressed condition was lower (58%) than expected from model predictions (67%). This could be due to several factors. First, it is unclear how the use of a novel word (*'the wet gorp'*) should influence contrastive inferences, compared to the use of a neutral pronoun (*'the wet one'*) as in Gelman and Markman. Some have argued that novel words have a special status compared to other linguistic components such as facts and that they prompt the formation of categories (Behrend, Scofield, & Kleinknecht, 2001). It is possible that children found the novel words more distracting and/or confusing than they would have a neutral pronoun like *one*.

This brings us to a second concern with regard to the assumption of overinformativenes which forms the basis of the contrastive inference. Indeed, if the inference is produced through reasoning about alternatives, the listener has to assume that the speaker is unaware of the listener's lack of knowledge about the meaning of the novel word. In this scenario, the listener can reason that if the speaker had meant to refer to the modified single alien (i.e., the wet alien which had no dry counterpart), they could simply have used that other alien's name (e.g., *'Touch the dax')*. In contrast, in a scenario where the speaker is aware that they are using unknown words, the same inference could not be derived (i.e., the speaker could not have simply used *'the dax'* since it would have been uninformative for the child). The issue with this assumption is that it makes the same use of modifier overinformative in the control trials (asking for *'the wet gorp'* when there was only one modified object in the array, e.g., one wet alien). This could have undermined the strength of the pragmatic inference in the critical trials, as the speaker would have been seen as relatively unreliable in their level of informativeness (Grodner & Sedivy, 2011). This is particularly likely given that only about half of the filler instructions in Gelman and Markman (1985) included an adjective and thus an overinformative expression (e.g., *'the red one'* when there is only one red object) while the other half used a simple noun (e.g., *'Show me the chair'*), thereby reducing the level of overinformativeness and 'unreliability' of the speaker (cf. Pogue, Kurumada, & Tanenhaus, 2016).

This contextual effect probably also explains why performance in the non-stressed condition was significantly lower in adults (and at chance in children): as demonstrated in Weber et al. (2006), the presence of stressed adjectives in the discourse context tends to reduce or inhibit contrastive inference for non-stressed adjectives (which do not appear as contrastive in comparison).

A final possibility mentioned a posteriori by some adult participants in our study was that the modifications (e.g., drops of water) partially hid some of the characters and made it difficult to perceive the difference between them, which could in some trials have led to the impression that both modified objects were part of a set (e.g., that both wet aliens had a dry counterpart). This would then have led to chance performance as it would have resulted in the use of the modifier being equally informative for both modified objects.

In conclusion, the predictions of the model were overall very close to the quantitative results obtained, however some factors such as stress and speaker reliability impacted the reliability of the inference. Next we briefly describe how such a discussion is not warranted in the case of our second experiment, precisely for the same reasons, we hypothesise, which lead to the difference between bilingual and monolingual groups to appear in one experiment but not in the other.

6.3.2 Experiment 2

The virtual impossibility of giving quantitative predictions for the results of Experiment 2 in a Bayesian framework acts as a further confirmation that this experiment is different in nature from our other tasks. Indeed, as we said before, performance relied entirely on the ability to recognise one emotional affect and pair it with the appropriate referent, without having to reason about speaker's intentions. In this way, the task was in fact akin to recognising the semantic meaning of a word and pairing it with an appropriate image (e.g., *'Show me the picture where he is happy'*). If an inference took place, it was related to world knowledge about how likely each modification (e.g., flower or star, dirt or breakage) was to have caused happy or sad emotions. Thus, this was in a way a *semantic* task which relied on knowing the *meaning* of each type of affect, whereas in our other experiments, knowledge of semantic meaning was either not sufficient (e.g., knowing the meaning of *'wet'* in Experiment 1, or *'kitten'* in Experiment 2) or not needed (e.g., for interpreting prosodic stress in Experiment 1, or for performing the correct inferences in Experiments 5 and 6).

We next turn to the results of our third and fourth experiments.

6.3.3 Experiments 3a, 3b and 4

In these experiments, participants had to infer the meaning of a novel word on the basis of its frequency in context. Thus, when a speaker identified a referent using a novel word (e.g., singling out a character and exclaming *'Oh, a kitten with a dax!'*), participants were expected to infer that the novel word referred to a feature of the referent which was informative, that is, rare enough to be remarked on. For example, if three kittens entered a room and a speaker pointed at one who had a bow and flower, the listener could use the frequency of each feature to choose between them as extension for the novel word. If all kittens had a bow and flower, the listener should be at chance between the two features (ratio between the two features=1). If the three kittens had a bow but only two had a flower, the flower would be a slightly more likely candidate (ratio between the two features=2/3), more so if two kittens had a bow but only the target one had a flower (ratio between the two features=1/2), and yet more if all three kittens had a bow but only the target one had a flower (ratio between the two features=1/3).

Once again, we can model this inference using Bayes' rule

$$P(f|w) = \frac{P(f)P(w|f)}{\sum_{f' \in F} P(f')P(w|f')} \tag{6.6}$$

Where f is the referent's feature and w the word used. Given that the novel word (e.g., *'dax'*) is unknown, it is by itself uninformative with regard to the referent's features being mentioned, which leads to w simply representing the utterance of *any* word. In other words, P($f|w$) is the probability of a certain feature being referred to, given a word has been uttered to describe its referent. If we assume that all features have an equal prior probability of being referred to, this is simply $P(w|f)$, or the likelihood of a word having been used to refer to a certain feature, which we can compute (according to the *Size Principle*, Tenenbaum & Griffiths, 2001) as inversely proportional to the ratio of the extension of this feature against the others

$$P(w|f) = \frac{|f|^{-1}}{\sum_{f' \in F} |f'|^{-1}} \tag{6.7}$$

These predictions were tested with adults online and with children in a laboratory setting by Frank and Goodman (2014). They found that adults's predictions were closely matched to the predictions made by a parameter-free model based on the equation above in four different conditions corresponding to the 'kitten' cases we outlined above. Thus, with equal ratio (e.g., one bow and one flower, or in their case one headband and one bandanna)

$$P(word|headband) = \frac{|headband|^{-1}}{|headband|^{-1} + |bandanna|^{-1}} = \frac{\frac{1}{1}}{\frac{1}{1} + \frac{1}{1}} = \frac{1}{2} \tag{6.8}$$

With a 1/2 ratio (e.g., one headband and two bandannas)

$$P(word|headband) = \frac{|headband|^{-1}}{|headband|^{-1} + |bandanna|^{-1}} = \frac{\frac{1}{1}}{\frac{1}{1} + \frac{1}{2}} = \frac{2}{3} \tag{6.9}$$

With a 2/3 ratio (e.g., two headbands and three bandannas)

$$P(word|headband) = \frac{|headband|^{-1}}{|headband|^{-1} + |bandanna|^{-1}} = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{3}} = \frac{3}{5} \tag{6.10}$$

And with a 1/3 ratio (e.g., one headband and three bandannas)

$$P(word|headband) = \frac{|headband|^{-1}}{|headband|^{-1} + |bandanna|^{-1}} = \frac{\frac{1}{1}}{\frac{1}{1} + \frac{1}{3}} = \frac{3}{4} \tag{6.11}$$

Agreggate performances in adults (responses were given in the form of quantitative 'bets' as to which feature was being referred to) in all four conditions were as follows: m=0.50, sd=0.14 (model prediction: 0.5), m=0.56, sd=0.23 (model prediction: 0.60), m=0.67, sd=0.28 (model prediction: 0.67) and m=0.70, sd=0.19 (model prediction: 0.75), the three latter significantly above chance. Children were tested in a simplified version of the second case (ratio=1/2) and found, in a first experiment, to actually perform *above* model predictions (m=0.81, sd=0.39 for 3-4 years old and m=0.88, sd=0.33 for 4-5 years old, model prediction: 0.67). In a first follow up experiment, the authors used a different construction (*'Here is a dinosaur with a dax'* instead of *'This is a dinosaur with a dax'*) to reduce the likelihood that the phrase would be stressed and/or interpreted contrastively and investigate if the results were specific to this particular construction. They found slightly lower, but still above chance performance (m=0.69, sd=0.47 for 3-4 years old and m=0.69, sd=0.47 for 4-5 years old). In a second follow up experiment, they investigated whether the salience of the unique object could have driven performance in the first two experiments by using a 'non-linguistic salience' condition where no novel words were used, and found that this did not seem to be the case.

In experiment 3a we attempted to replicate these results in adults and children of similar age (4 to 6 years old) by using a slightly modified version of Frank and Goodman's (hereafter F & G) paradigm implemented on a touchscreen laptop. To our surprise, we found performance to be significantly *below* chance in children and at chance in adults. This was hypothesised as the result of uncertainty introduced regarding the target character, which was singled out by a red circle in our case whereas it was singled out by a more unambiguous point in F & G. This uncertainty had made available an alternative salient answer which adults and children had taken up on in similar proportions. However, examination of answer patterns revealed that the proportion of consistent target answers (across all four trials) in adults was about 40%, whereas they were absent in children. To further investigate this, we conducted a follow up experiment (Experiment 3b) where the target character was made explicit (*'The kitten circled in red'*). Here again, adults produced the target answer, this time according to the model's quantitative predictions (as it was not balanced by alternative answers): i.e., 66%, but children did not seem to derive the target pragmatic inference and performed at chance.

One crucial difference between F & G's study and ours is that, to make computer implementation easier and reduce as much as possible the memory component of the task, we displayed both training and test items on the same screen *simultaneously*. This effectively made our display a version of the 2/3 ratio condition in F & G (two instances of target object, e.g., blue novel object and three of non-target, e.g., pink novel object). This condition, however, is set apart from the other ones. Indeed, it is the only one in F & G's where the target object is *not* unique in the display. This makes this case the only one where the frequency inference cannot be substituted by an easier contrastive or exclusion inference. Indeed, in the other cases it is possible to interpret the phrase as *'This is a dinosaur with a dax* [and these other ones are not]'. Thus, it is an open question to which extent the low level of adult performance in the 2/3 condition (0.56) was simply due to the extension ratio being highest (2/3 vs. 1/2 or 1/3) or was also influenced by the unavailability of this contrastive reading. One way of verifying this would be to organise a display where the ratio is constant but the target object is not unique (e.g., 2 headbands and 4 bandannas). However, this introduces another variable which might impact performance, namely the visual complexity of the display.

In view of this, we would like to suggest that the reason children succeeded in F & G but failed to derive the target frequency inference in our experiment is that they were actually performing an exclusion inference (*'That was a dinosaur with a dax so the other one was not'*), which was not warranted in our display since the target object in it was not unique to the target character. This is in line with recent work showing that children are able to perform 'simpler' exclusion inferences but fail to display asymmetries in answers which would indicate full reasoning on the basis of informativeness in scalar implicature paradigms (J. Sullivan et al., 2017). This would also explain the relatively high levels of performance, as this type of reasoning functions by definition in an 'all-or-nothing' rather than graded manner. The authors did attempt to check for a contrastive or uniqueness effect in their follow up experiments. However, we would argue that the use of a locative (*'Here is a dinosaur with a dax'*) does not completely rule out a contrastive or exclusion explanation. Similarly, the 'non-linguistic salience' condition of the second follow-up experiment does not rule out a preference for the unique object, as the training and test question were simply phrased as *'Here is a dinosaur'* and *'Can you find another one?*, which arguably would lead children to believe that the referent of 'another one' is 'dinosaur', and to perform at chance. Ruling out an effect of uniqueness of target could only be achieved by using a display where the target object was not unique and finding that children still performed the expected inference (as adults do). Contrary to this, we found in our experiment that this display seemed to prevent children from performing the target pragmatic inference.

We now turn to the results in our fifth experiment, which investigated level of extension of a category based on presentation of exemplars.

6.3.4 Experiment 5b

In our fifth experiment we used a paradigm similar to Xu and Tenenbaum (2007a) to investigate sensitivity to sampling in extension of novel categories. These authors propose a Bayesian model for word learning which combines prior beliefs and given evidence to yield a probability distribution over the possible extensions of the new word. Since this model is inferential in nature and involves reasoning, it makes different predictions than a purely associative model regarding how information about the source of the evidence should be treated by learners. In an associative model of word learning (where word extensions are learned cross-situationally through simultaneous exposure to word and referent), there should be no effect of sampling and only number of labelled referents should have an influence over subsequent extensions by strengthening the associative relation between word and referent. E.g., In the absence of negative feedback or information to the contrary, a dalmatian being labelled 'dax' twice should drive learners to the same conclusions regardless of the process which led to the labelling. In a Bayesian framework, on the other hand, exemplar choices performed by a knowledgeable teacher with ostensive educational purposes should be regarded as informative, whereas choices performed by an ignorant learner with a different goal should be seen as uninformative and dismissed from the reasoning process. In their study, Xu and Tenenbaum (2007a) presented children and adults with a picture showing sets of novel objects which could be grouped into broad 'basic-level' categories (similar level as dogs) based on shape, and narrower 'subordinate' categories (similar level as dalmatians). In the teacher-driven condition, the experimenter pointed at three objects from the same subordinate category in turn and labelled them with the same novel word, e.g.,*'blicket'*. In the learner condition, the experimenter pointed at and labelled one subordinate exemplar and asked the learner to label two other exemplars on the picture themselves, promising a reward if they 'got it right', which prompted conservatism instead of exploration in learners' choices, and insured that the exemplars chosen would be of the same subordinate category (regardless of participant's beliefs about the true extension of the novel word). The test of extension then consisted in the experimenter pointing a five different objects in turn, including objects from the same subordinate category, different subordinates from the same basic-level category and objects from a different basic-level category, and asking the participant *'Is this a blicket?*

Our adaptation of Xu and Tenenbaum's experiment provided us with an opportunity to test the predictions of their model in a slightly modified paradigm. Indeed, in our version we provided participants with only two exemplars of the target subordinate category to avoid potential ceiling effects. The exemplars were also presented separately rather than singled out in a picture containing all objects, to increase retention in short term memory. This is because our experiment was implemented as a computerised task which meant the screen would switch between training and testing.

Formally, the Bayesian inference can be written as

$$P(m|e) = \frac{P(m)P(e|m)}{\sum_{m' \in M} P(m')P(e|m')} \tag{6.12}$$

Where m is the meaning hypothesis (or possible extension of the word: 'subordinate level' or 'basic level') and e is the evidence or exemplars provided (e.g., one labelled examples, two labelled examples, etc.). $P(m)$ constitutes the prior belief about extension, or tendency to extend the novel label to the subordinate or basic-level prior to seeing any exemplars, whereas $P(e|m)$ is the likelihood of having sampled the exemplars given a specific meaning or extension level.

As previously described, Xu and Tenenbaum's model assumes taxonomic categories where the probability of sampling from each level is proportional to the 'height' of its branch, i.e., the higher the level, the more likely it is to sample from it as it is broader (e.g., animals are more likely than dogs which are more likely than dalmatians). In this particular paradigm, the extension of each category is approximated as the number of exemplars seen. The *Size Principle* which follows from this reflects the intuition of a 'suspicious coincidence' whereby exemplars from a lower-level category have a higher chance of having been sampled from this category only than from a higher-level category (in which case you would expect more variety/other subordinates). This 'suspicion' increases exponentially with each additional exemplar. Thus, for each exemplar the sampling probability is the inverse of the category extension, and the likelihood is computed as their product

$$P(e|m) = \prod_{j=1}^{n} P(e_j|m) \tag{6.13}$$

Where n is the number of exemplars. In the teacher condition, where every exemplar is considered informative we have

$$P(e_1...e_n|m) = \left(\frac{1}{|m|}\right)^n \tag{6.14}$$

In the learner-driven condition, on the other hand, only the first exemplar (given by the knowledgeable teacher) is informative regarding the true meaning distribution, thus in effect equating a likelihood where n=1. As we have seen before, the prior over extension type is not uniform between subordinate and basic-level choices (i.e., there is a bias towards basic level choices). Given this and the need for all meaning hypotheses to sum to one, the authors fit as prior a beta distribution uniquely described by one parameter β=5 whereby all things being equal, basic level hypotheses are believed to be five times as likely as subordinate hypotheses (e.g., 0.17 vs. 0.83). In Xu and Tenenbaum's (2007a) paradigm (hereafter X & T), each of the two basic level categories has three subordinate categories which each contain five objects. Thus in the teacher condition the likelihood for subordinate level after three exemplars from the same subordinate category have been labelled is $(1/5)^3$=0.008, which combined with prior yields a posterior probability of 0.84. In the learner condition, since only the first exemplar is informative, the likelihood for subordinate level is 1/5=0.2, which combined with prior yields a posterior probability of 0.37. The numbers obtained by X & T were slightly lower in children (mean age 4;0, 71% subordinate choices in teacher condition and 29% in learner condition) and slightly higher in adults for the teacher condition (92% subordinate choices in teacher condition and 35% in learner condition). This is interesting as it confirms a previous tendency found in (Xu & Tenenbaum, 2007b) whereby children appeared *less* likely than adults to display a basic-level bias.

$$P(sub|threesubs, teacher) = \frac{P(twosubs|sub)P(sub)}{P(twosubs)} = \frac{0.008 * 0.17}{0.008 * 0.17 + 00.0002963 * 0.83} = 0.84 \quad (6.15)$$

$$P(sub|onesub, learner) = \frac{P(twosubs|sub)P(sub)}{P(twosubs)} = \frac{0.2 * 0.17}{0.2 * 0.17 + 0.0667 * 0.83} = 0.37 \quad (6.16)$$

In our paradigm, there were only six objects which could be seen simultaneously at test time: two target subordinates, two other subordinates (one of each) from the same basic level and two objects from another basic level category. Thus the extension for target basic level was 4 objects (against 15 in X & T) and the extension of the subordinate category was 2 objects (against 5 in X & T). If we conserve a β of value 5 (which is appropriate given our number of basic and subordinate categories is approximately equal to that of X & T's, i.e., 2 basic levels and 5 subordinate categories in total, whereas they had 6 subordinate categories, three in each basic level). This means in our teacher condition the likelihood for subordinate level should be $(1/2)^2$=0.25, which combined with prior yields a posterior probability of 0.44, whereas in the learner condition the likelihood is 1/2 which combined with prior yields 0.28.

$$P(sub|twosubs, teacher) = \frac{P(twosubs|sub)P(sub)}{P(twosubs)} = \frac{0.25 * 0.17}{0.25 * 0.17 + 0.0625 * 0.83} = 0.44 \tag{6.17}$$

$$P(sub|onesub, learner) = \frac{P(onesub|sub)P(sub)}{P(onesub)} = \frac{0.50 * 0.17}{0.50 * 0.17 + 0.25 * 0.83} = 0.28 \tag{6.18}$$

This proportion of subordinate choices is lower than the numbers we obtained in children in both the teacher and learner conditions (72% and 42% respectively). This could be due to several reasons. First, as we said before, X & T themselves acknowledged a lesser basic-level bias in children, which means the β parameter which accounts for this preference might be better set at a lower value when modelling children's choices compared to adults. Another reason why β should perhaps be set lower than the (relatively high) value of 5 in our paradigm, especially in the teacher condition, is that (unlike in X & T, and in our learner condition) the two exemplars were presented simultaneously and separately from the other objects, thus increasing the salience of the subordinate category. It is also worth noting that even in the test part of the trials, the objects were not 'bundled' together as in X & T and we had a lesser number of each type, which might have lowered the salience of the basic category. It is an interesting question whether the type/token ratio would have any influence on prior biases or choices after seeing exemplars. Finally, it is likely that a basic level bias would be impacted by the level of similarity between subordinate categories, i.e., the more similar they are, the more likely they are to belong to the same basic level and thus the more likely a participant will be to give a basic level answer. It is possible that our subordinate categories differed from each other to a greater extent than that of X & T, or that, as we mentioned before, the grouping of a high number of objects from the different subordinate categories together made their similarities more salient.

We would also like to suggest the possibility that participants might have reasoned in a simpler way, i.e., they might have considered that providing two exemplars from the same subordinate category was twice as ambiguous evidence as different subordinates would have been. Indeed, two different subordinates would have unequivocally lead to learning the word for the basic level category, whereas two same subordinates could mean either the subordinate or basic level category. If reasoning by exclusion, participants might have simply thought that if the teacher had wanted to teach them the basic level category, they would have used different subordinates, similarly to the classic some-but-not-all scalar implicature (cf. Figure 6.1). This type of reasoning could then have boosted the subordinate choices in teacher condition, either because they relied on exclusion 'all-or-nothing' reasoning, or because they reasoned that getting two same subordinates was twice as ambiguous as getting two different ones, and thus twice less likely to be used for basic level category teaching (i.e., 67% vs. 33%, which numbers are closer to the ones we obtained).

In conclusion, while the model in Xu and Tenenbaum (2007b) does provide explanatory power, it is delicate to give definitive values for free parameters, such as the beta parameter for basic-level bias, which they set at the value of 5 in their study, a value which might be too high in a different paradigm such as ours, where the similarity between basic level objects is not as salient.

6.3.5 Experiment 6b

With our last experiment we enter unknown territory in terms of quantitative predictions. Indeed, to our knowledge this type of pragmatic inference has not previously been modelled using a Bayesian framework. In this experiment, participants witnessed a 'teacher' perform a novel action with a novel object while uttering a novel word. In a control condition there was no previous preparation, whereas in the 'action-emphasised' condition, this was preceded by the teacher performing *another* action with the novel object before performing the target action and uttering the novel word. This emphasised that the novel action was a more likely referent for the novel word, since it was more recent to the common ground (i.e., if the novel object was the intended referent, why would the speaker not have labelled it immediately?).

Previous studies have shown that young children are able to infer that the referent of a novel word is the most novel referent in the common ground between speaker and listener. In Akhtar et al. (1996), two-year olds were given three novel, nameless objects. Both experimenter and child played with them and performed several actions on them. The three objects were then placed in a clear plastic box along with a fourth, yet unseen object, also novel and nameless. The experimenter then uttered a novel word enthusiastically (five times in total, given that parents were instructed to produce similar utterances) while looking at the box with objects (e.g., *'Look,I see a modi! A modi! I see a modi in there!'*). After this, participants played again with all four objects to ensure equal level of familiarity and exposure. At test, children were first asked to choose between the four objects (comprehension test) (e.g., *'Show me the modi'*). Half of the children (n=8) successfully picked the target object. In addition, five children produced the novel word appropriately, either spontaneously or through elicited production (production test). Akthar et al. then concluded that 10 children in total (62.5%) could be considered as having learned and correctly extended the novel word.

In another experiment similar to ours, Tomasello and Akhtar (1995) examined whether this ability to use discourse novelty for word learning extended to being able to distinguish between novel object and novel action as the referent of a novel word. To this end, they constructed an 'object novel' condition and an 'action novel' condition. In the object novel condition, child and experimenter performed the target action with multiple novel objects for a length of two minutes before performing the target action with the target object five times and uttering the novel word, then repeating the full sequence (a total of ten utterances of the novel word). In the action novel condition, child and experimenter performed multiple actions with the target object before performing the target action with the target object five times and uttering the novel word, then repeating the full sequence (again for a total of ten utterances of the novel word). Thus the last sequence was the same in both conditions, whereas the play and preparations that had led to it were different. Crucially, there were no morphosyntactic markers (no article or tense) as to which part of speech (noun or verb) the novel word belonged to. Here again, children were credited with having learned the novel word either through comprehension (they were asked *'Look over there! Can you show me modi?'* by the experimenter while designating the area of the room where training had taken place) or through spontaneous production. Children who picked or performed another action with the target object were credited with an object response, while children children who performed the target action with another object were credited with an action response. Children who performed target action with target object (n=2 in the action novel condition) were credited with action responses on the basis of their performance at pre-test, where they had never produced an action in reponse to a prompt for a familiar object. In total, 9 out of 12 children (75%)

produced an action answer in the action novel condition, against 3 (25%) in the object novel condition. Tomasello and Akhtar (1995) then concluded that two year olds were reliably able to use pragmatic inferences based on discourse novelty of a referent to learn and extend the meaning of a novel word to either action or object.

We can tentatively try and model the inference using Bayes' rule as

$$P(r|w) = \frac{P(p)P(w|r)}{\sum_{r' \in R} P(r')P(w|r')} \tag{6.19}$$

Where r stands for a specific referent or part of speech (action or object, verb or noun). Thus, P($r|w$) is the posterior probability of a referent given a novel word has been uttered, P(r) is the prior probability of the referent or part of speech and P($w|r$) is the likelihood of a word being given *at a certain time* for a specific referent.

This inference is, in fact, also a type of frequency inference (cf. Experiments 3a and 3b) where frequency is expressed in terms of novelty and here again, the novel word by itself is uninformative. However, because novelty is a *temporal* measure, inference is based this time on the *timing* of the novel word utterance relative to the introduction of the referent. Thus, P($w|r$) is the likelihood of a label being provided at a certain time for a referent, which is proportional to the novelty of this referent, or inversely proportional to the frequency of the referent in context at the time of the labelling event (how much it has been seen before the novel word is given).

The difficulty here resides in defining a quantitative measure of novelty for each referent. This is more difficult than for the frequency inference case in Experiments 3a and 3b, where objects can be counted, as they are discrete units. Time, on the other hand, is by nature a continuous and relative measure. A simplifying assumption could be to count as a discrete unit each event of performing an action with an object and record the occurrence of each referent across all events in the relevant context. Thus, in the action emphasised condition of our paradigm we would have an extension of

1/2 for the target action (which appeared in one out of two events) and 2/2 for the novel object (which appeared in both events) and the likelihood for action would be computed as

$$P(w|action) = \frac{|r|^{-1}}{\sum_{r' \in R} |r'|^{-1}} = \frac{|action|^{-1}}{|action|^{-1} + |object|^{-1}} = \frac{\frac{2}{1}}{\frac{2}{1} + \frac{2}{2}} = \frac{2}{3} \tag{6.20}$$

However, such a measure is problematic for at least two reasons. First, this measure does not take into account exposure to an object when no actions were performed, which should still lower the novelty of the object. Second, it attributes the same weight to all events involving a referent after it has been introduced, whereas it could be expected that the exact number of events performed, or length of time elapsed since introducing the referent, matters less than the fact that one introduction took place simultaneously with, or much closer to, the utterance of the novel word.

Another way of measuring novelty could be simply to quantify in units of time, e.g., minutes, the fraction of the time occupied by each referent between the introduction of the first referent and the labelling event or utterance of the novel word. However, if we consider a 10 minute-training where the second referent had been introduced 1 minute before the label (which already seems a longer period than in the experiments described), this would yield a likelihood of 0.91, a much higher number than the one found in Tomasello and Akhtar (1995) (hereafter T & A). For a simultaneous introduction of word and referent (and thus an exposure of, or close to 0), which was the case in our paradigm, we would actually have a likelihood of, or close to 1, whereas the results we obtained were much lower and only above chance in the bilingual group (0.57).

Despite this, it is likely that the difference between ours and T & A in terms of length of exposure contributed to the quantitative differences found in the results, despite the older age of our participants. Precisely with, in our mind, the goal of being able to distinguish and quantify exposure to each referent, ours was a very stripped down and brief experimental manipulation, with each training video lasting about ten seconds. This means that participants had less time to form and retain the mapping between label and referent, compared to T & A where the sequence and labelling were repeated a total of ten times in each condition, which helped strengthen the result of the pragmatic inference made from the previously seen preparations. In addition, two competing effects are at play in terms of salience resulting from frequency. While it is true that more rare (or more novel) objects and properties are more salient, it is also true that the additional exposure resulting from greater frequency will lead to objects or properties in high numbers being also salient and more lastingly retained (Tarenskeen, Broersma, & Geurts, 2015). Think about entering a living room which is entirely filled with blue furniture and objects. You attention will probably be attracted by the one red cushion on the sofa, but equally the colour blue is likely to be salient to you and memorised even after leaving the room. In this view, when relatively short periods of time are involved (as in our videos), the salience from length of exposure is likely to be competing with the effect of the novelty-based pragmatic inference, resulting in performances being closer to chance. Similarly, we were surprised to find that performance in the neutral control condition was at chance between action and object, while previous research and the 'whole object constraint' (e.g., Markman, 1990) would have led us to expect a bias towards mapping a novel word to a novel object. However, once again, it is possible that the manipulations of the object in our action-emphasised condition also had the effect of attracting further attention to the object, thus partially cancelling the effect of the novelty inference.

In conclusion, computational models of pragmatic inference, cross-situational word learning and exemplar learning which use Bayesian inference prove to be useful ways of formalising these processes and providing quantitative predictions. However, when dealing with behavioural data, quantitative outcomes are sensitive to a variety of factors which can have a substantial impact on performance. Going back to the puzzle of the bilingual lexicon outlined in the introduction, and the poor performance of the model when faced with multilingual input, we would like to suggest, in light of our findings, that a Bayesian model of cue integration for bilingual word learning needs to be adapted with more flexibility in combining cues, i.e., less weight from prior 'default' biases towards similarity matches, whole objects or basic level, or less tendency to 'jump to conclusions', and a higher willingness to let go of these priors in order to consider pragmatic cues, as well as a more optimally rational behaviour in deriving inferences from these cues.

6.4 Future directions

As with all works which are limited in time, this thesis leaves open some avenues for further research, along with unanswered questions.

One concerns the exact part played by the pragmatic abilities described here in previously found advantages in word learning paradigms in the work of Yow and colleagues. Indeed, while our work demonstrates that such advantages could not have been *solely* due to better inhibitory control or attentional biases towards social cues, we do not know to which extent the abilities we examine here contributed to these results *in addition* to other potential differences. Further research will thus have to investigate whether these different skills can be examined separately and whether a specific developmental timeline can be established for each of them. This should then help us have a better understanding of pragmatic mechanisms for fast mapping and word learning in general in allowing us to model the flexible process of cue integration whereby individuals will use and rely on different types of cues depending on their experience.

In this view, it would also be useful to investigate the domains where cue integration is performed in a graded 'Bayesian' manner from those where reasoning happens in a more exclusive, 'all-or-nothing' fashion. As we have seen, exclusion reasoning seems to be more prevalent in children, who might use it as a stepping stone for acquiring a full pragmatic competence. However, the type of evidence received can also influence the reasoning process, as we have seen in the case of hierarchical category extensions, which are sometimes performed in a graded manner (e.g., when one exemplar is given) and sometimes performed in a more extreme manner (e.g., when multiple exemplars are given).

In the course of this research, we were sometimes faced with children's inability to derive pragmatic inferences which adults could succeed at. While we provided a number of hypotheses for why this might have been the case, relying on both the developmental literature (e.g., recent work showing that children need a higher number of, or more reliable cues for performing certain types of inference) and on follow up experiments where we further explored some open questions (e.g., distinguishing between frequency and exclusion inference in Experiment 3b, or showing that children could succeed in using a single pragmatic cue in an ostensive context), there is more than enough room for future research to examine the factors, including methodological, which impact successful performance of pragmatic inference in certain paradigms.

All in all, lazy humans appear to engage in pragmatic reasoning when they have an incentive to do so (it is necessary) or are more used to doing it (it has become more default). Since bilingually exposed individuals check both boxes, it stands to reason that they should demonstrate it more prevalently.

References

Abbot-Smith, K., Imai, M., Durrant, S., & Nurmsoo, E. (2017). The role of timing and prototypical causality on how preschoolers fast-map novel verb meanings. *First Language*, *37*(2), 186–204.

Adams-Bedford, J., Wallis, G., & Backus, B. (2013). The impact of intention, action, and learnt contingency on visual perception. *Journal of Vision*, *13*(9), 754–754.

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, *67*(2), 635–645.

Akhtar, N., Menjivar, J., Hoicka, E., & Sabbagh, M. A. (2012). Learning foreign labels from a foreign speaker: The role of (limited) exposure to a second language. *Journal of child language*, *39*(5), 1135–1149.

Alcántara, J. I., Weisblatt, E. J., Moore, B. C., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *45*(6), 1107–1114.

Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press.

Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of memory and language*, *60*(2), 270–290.

Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of memory and language*, *53*(1), 60–80.

Andrés-Roqueta, C., & Katsos, N. (2017). The contribution of grammar, vocabulary and theory of mind in pragmatic language competence in children with autistic spectrum disorders. *Frontiers in psychology*, *8*, 996.

Antoniou, K., & Katsos, N. (2017). The effect of childhood multilingualism and bilectalism on implicature understanding. *Applied Psycholinguistics*, *38*(4), 787–833.

Antoniou, K., Veenstra, A., Kissine, M., & Katsos, N. (2018). The impact of childhood bilingualism and bi-dialectalism on pragmatic interpretation and processing.

AuerJr, E. T., & Bernstein, L. E. (2007). Enhanced visual speech perception in individuals with early-onset hearing impairment. *Journal of Speech, Language, and Hearing Research*.

Bach, K. (1994). Conversational impliciture. *Mind & language*, *9*(2), 124–162.

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental psychology*, *29*(5), 832.

Barac, R., & Bialystok, E. (2011). Cognitive development of bilingual children. *Language Teaching*, *44*(1), 36–54.

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in

early language development. *Cognitive psychology*, *60*(1), 40–62.
Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.
Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind.* MIT press.
Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in psychology*, *2*, 324.
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
Bates, E., MacWhinney, B., & MacWhinney, B. (1987). Competition, variation, and language learning. *Mechanisms of language acquisition*, 157–193.
Behrend, D. A., Scofield, J., & Kleinknecht, E. E. (2001). Beyond fast mapping: Young children's extensions of novel words and novel facts. *Developmental Psychology*, *37*(5), 698.
Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600–1610.
Ben-Zeev, S. (1977). The influence of bilingualism on cognitive strategy and cognitive development. *Child development*, 1009–1018.
Berman, J. M., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, *107*(2), 87–99.
Berman, J. M., Graham, S. A., Callaway, D., & Chambers, C. G. (2013). Preschoolers use emotion in speech to learn new words. *Child development*, *84*(5), 1791–1805.
Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental psychology*, *24*(4), 560.
Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*(1), 67–73.
Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development*, *11*(4), 485–508.
Bialystok, E., & Craik, F. I. (2010). Cognitive and linguistic processing in the bilingual mind. *Current directions in psychological science*, *19*(1), 19–23.
Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531.
Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition*, *140*, 122–134.
Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39–53.
Bloom, P. (2000). *How children learn the meanings of words: Learning, development and conceptual change.* MIT Press.[aJRH].
Bohn, M., Tessler, M. H., & Frank, M. C. (2019). Integrating common ground and informativeness in pragmatic word learning.
Bornstein, M. H. (1985). Colour-name versus shape-name learning in young children. *Journal of Child Language*, *12*(2), 387–393.
Bruner, J. (1985). Child's talk: Learning to use language. *Child Language Teaching*

and Therapy, *1*(1), 111–114.
Byers-Heinlein, K. (2017). Bilingualism affects 9-month-old infants' expectations about how words refer to kinds. *Developmental science*, *20*(1), e12486.
Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: insights from bilingual infants. *Developmental Psychobiology*, *56*(2), 274–291.
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.
Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*(3), 407–416.
Caillies, S., & Butcher, K. (2007). Processing of idiomatic expressions: Evidence for a new hybrid view. *Metaphor and Symbol*, *22*(1), 79–108.
Caillies, S., & Declercq, C. (2011). Kill the song—steal the show: what does distinguish predicative metaphors from decomposable idioms? *Journal of psycholinguistic research*, *40*(3), 205–223.
Callanan, M. A., Repp, A. M., McCarthy, M. G., & Latzke, M. A. (1994). Children's hypotheses about word learning constraints: is there a basic-level constraint? *Journal of Experimental Child Psychology*, *57*(1), 108–138.
Cane, J. E., Ferguson, H. J., & Apperly, I. A. (2017). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 591.
Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant behavior and development*, *21*(2), 315–330.
Carston, R. (1995). Quantity maxims and generalised implicature. *Lingua*, *96*(4), 213–244.
Carston, R. (2004). Relevance theory and the saying/implicating distinction. *The handbook of pragmatics*, 633–656.
Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*(2), 159–199.
Champoux-Larsson, M.-F., & Dylman, A. S. (2018). A prosodic bias, not an advantage, in bilinguals' interpretation of emotional prosody. *Bilingualism: Language and Cognition*, 1–9.
Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, *25*(5), 975–979.
Chomsky, N. (1980). Rules and representations. *Behavioral and brain sciences*, *3*(1), 1–15.
Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language*, *15*(2), 317–335.
Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition*, *64*(1), 1–37.
Clark, E. V. (2009). *First language acquisition.* Cambridge University Press.
Clark, E. V., & Grossman, J. B. (1998). Pragmatic directions and children's word learning. *Journal of Child Language*, *25*(1), 1–18.
Cohen, L. J. (1971). Some remarks on grice's views about the logical particles of natural language. In *Pragmatics of natural languages* (pp. 50–68). Springer.

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of memory and language*, *45*(4), 721–736.
Colunga, E., Brojde, C., & Ahmed, S. (2012). Bilingual and monolingual children attend to different cues when learning new words. *Frontiers in Psychology*, *3*, 155.
Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological review*, *112*(2), 347.
Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and language*, *41*(3), 365–397.
Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language*, *12*(3), 643–661.
Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148–153.
Cummins, J., & Mulcahy, R. (1978). Orientation to language in ukrainian-english bilingual children. *Child development*, 1239–1242.
Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 892.
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*(1), 25–45.
Davies, C., & Katsos, N. (2009). Are interlocutors as sensitive to over-informativeness as they are to under-informativeness? proceedings of the workshop on production of referring expressions: Bridging computational and psycholinguistic approaches.
Davies, C., & Katsos, N. (2013). Are speakers and listeners 'only moderately gricean'? an empirical response to engelhardt et al.(2006). *Journal of Pragmatics*, *49*(1), 78–106.
Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12-and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental psychology*, *36*(4), 511.
Degen, J., Franke, M., & Jager, G. (2013). Cost-based pragmatic inference about referential expressions. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).
Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science*, *39*(4), 667–710.
De Heering, A., Rossion, B., & Maurer, D. (2012). Developmental changes in face recognition during childhood: Evidence from upright and inverted faces. *Cognitive Development*, *27*(1), 17–27.
De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, *35*(6), 1189–1211.
Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: An empirical examination. *Developmental Psychology*, *41*(3), 451.
Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental psychology*, *37*(5), 630.
Dromi, E. (1999). Early lexical development. *The development of language*, *99131*.

Dunn, L. M., & Dunn, D. M. (2009). *The british picture vocabulary scale.* GL Assessment Limited.

D'Entremont, B., & Muir, D. (1999). Infant responses to adult happy and sad vocal and facial expressions during face-to-face interactions. *Infant Behavior and Development*, *22*(4), 527–539.

Elgier, A. M., Jakovcevic, A., Mustaca, A. E., & Bentosela, M. (2012). Pointing following in dogs: are simple or complex cognitive mechanisms involved? *Animal Cognition*, *15*(6), 1111–1119.

Elston-Güttler, K. E., & Williams, J. N. (2008). First language polysemy affects second language meaning interpretation: Evidence for activation of first language concepts during second language reading. *Second Language Research*, *24*(2), 167–187.

Engel de Abreu, P. M. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory*, *19*(5), 529–537.

Engelhardt, P. E., Bailey, K. G., & Ferreira, F. (2006). Do speakers and listeners observe the gricean maxim of quantity? *Journal of Memory and Language*, *54*(4), 554–573.

Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, *26*(7), 1090–1097.

Fazey, I., Fazey, J. A., Fischer, J., Sherren, K., Warren, J., Noss, R. F., & Dovers, S. R. (2007). Adaptive capacity and learning to learn as leverage for social–ecological resilience. *Frontiers in Ecology and the Environment*, *5*(7), 375–380.

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.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant behavior and development*, *8*(2), 181–195.

Frank, M. C. (2014). Learning words through probabilistic inferences about speakers' communicative intentions. *Language in Interaction: Studies in honor of Eve V. Clark*, *12*.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive psychology*, *75*, 80–96.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological science*, *20*(5), 578–585.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, *4*, 1–1.

Gathercole, V. C. M., & Moawad, R. A. (2010). Semantic interaction in early and late bilinguals: All words are not created equally. *Bilingualism: Language and Cognition*, *13*(4), 385–408.

Gelman, S. A., & Markman, E. M. (1985). Implicit contrast in adjectives vs. nouns: Implications for word-learning in preschoolers. *Journal of Child language*, *12*(1), 125–143.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Center for the Study of Reading Technical Report; no. 257*.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological science*, *17*(8), 684–691.

Gibbs Jr, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of memory and language*, *28*(5), 576–593.

Gibbs Jr, R. W., & O'Brien, J. E. (1990). Idioms and mental imagery: The metaphorical motivation for idiomatic meaning. *Cognition*, *36*(1), 35–68.

Giora, R. (2003). *On our mind: Salience, context, and figurative language*. Oxford University Press.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language learning and development*, *1*(1), 23–64.

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition*, *6*(1), 1–15.

Golinkoff, R. M., Jacquet, R. C., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles may underlie the learning of verbs. *Child development*, *67*(6), 3101–3119.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework [*]. *Journal of child language*, *21*(1), 125–155.

Gollan, T. H., Salmon, D. P., Montoya, R. I., & Galasko, D. R. (2011). Degree of bilingualism predicts age of diagnosis of alzheimer's disease in low-education but not in highly educated hispanics. *Neuropsychologia*, *49*(14), 3826–3830.

Goodman, N., Tenenbaum, J. B., & Black, M. J. (2008). A bayesian framework for cross-situational word-learning. In *Advances in neural information processing systems* (pp. 457–464).

Graf, C., Degen, J., Hawkins, R. X., & Goodman, N. D. (2016). Animal, dog, or dalmatian? level of abstraction in nominal referring expressions. In *Cogsci.*

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824.

Grassmann, S., & Tomasello, M. (2010). Prosodic stress on a word directs 24-month-olds' attention to a contextually new referent. *Journal of Pragmatics*, *42*(11), 3098–3105.

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

Green, D. W. (2011). Language control in different contexts: the behavioral ecology of bilingual speakers. *Frontiers in psychology*, *2*, 103.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530.

Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive development*, *28*(1), 41–50.

Grice, H. P. (1975). Logic and conversation. *1975*, 41–58.

Groba, A., De Houwer, A., Mehnert, J., Rossi, S., & Obrig, H. (2018). Bilingual and monolingual children process pragmatic cues differently when learning novel adjectives. *Bilingualism: Language and Cognition*, *21*(2), 384–402.

Grodner, D., & Sedivy, J. C. (2011). 10 the effect of speaker-specific information on pragmatic inferences. *The processing and acquisition of reference*, 239.

Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press.
Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23–B34.
Hanks, P. (2004). The syntagmatics of metaphor and idiom. *International Journal of Lexicography*, *17*(3), 245–274.
Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61.
Healey, E., & Skarabela, B. (2008). Are children willing to accept two labels for a single object?: A comparative study of mutual exclusivity in monolingual and bilingual children. *of the 2007*, 48.
Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836.
Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition*, *149*, 104–120.
Hemsley, G., Holm, A., & Dodd, B. (2013). Conceptual distance and word learning: Patterns of acquisition in samoan–english bilingual children. *Journal of child language*, *40*(4), 799–820.
Henderson, A. M., & Scott, J. C. (2015). She called that thing a mido, but should you call it a mido too? linguistic experience influences infants' expectations of conventionality. *Frontiers in Psychology*, *6*, 332.
Hirsh-Pasek, K., Michnick Golinkoff, R., & Hollich, G. (1999). Trends and transitions in language development: Looking for the missing piece. *Developmental neuropsychology*, *16*(2), 139–162.
Hobbs, G., & Vignoles, A. (2007). *Is free school meal status a valid proxy for socio-economic status (in schools research)?* Centre for the Economics of Education, London School of Economics and ....
Hochstein, L., Bale, A., & Barner, D. (2018). Scalar implicature in absence of epistemic reasoning? the case of autism spectrum disorder. *Language Learning and Development*, *14*(3), 224–240.
Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. *Meaning, form, and use in context: Linguistic applications*, *11*, 42.
Horn, L. R., & Ward, G. L. (2004). *The handbook of pragmatics*. Wiley Online Library.
Horowitz, A. C., & Frank, M. C. (2016). Children's pragmatic inferences as a route for learning about the world. *Child development*, *87*(3), 807–819.
Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The trouble with quantifiers: exploring children's deficits in scalar implicature. *Child development*, *89*(6), e572–e593.
Horowtiz, A., & Frank, M. C. (2014). Preschoolers infer contrast from adjectives if they can access lexical alternatives. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 36).
Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157.
Horst, J. S., Scott, E. J., & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science*, *13*(5), 706–713.
Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117.
Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes

the development of the mutual exclusivity bias. *Infancy*, *15*(2), 125–150.
Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology*, *58*(3), 376–415.
Imai, M., Haryu, E., Okada, H., Lianjing, L., & Shigematsu, J. (2006). 17 revisiting the noun-verb debate: A cross-linguistic comparison of novel noun and verb learning in english-, japanese-, and chinese-speaking children. *Action meets word: How children learn verbs*, 450.
Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in chinese-, english-, and japanese-speaking children. *Child development*, *79*(4), 979–1000.
Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in spanish-speaking infants and toddlers. *Journal of Child Language*, *20*(3), 523–549.
Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision research*, *39*(21), 3621–3629.
Jaswal, V. K., & Hansen, M. B. (2006). Learning words: Children disregard some pragmatic information that conflicts with mutual exclusivity. *Developmental science*, *9*(2), 158–165.
Jaszczolt, K. M. (2009). Cancelability and the primary/secondary meaning distinction. *Intercultural Pragmatics*, *6*(3), 259–289.
Junker, D. A., & Stockman, I. J. (2002). Expressive vocabulary of german-english bilingual toddlers. *American Journal of Speech-Language Pathology*, *11*(4), 381–394.
Kahneman, D., & Egan, P. (2011). *Thinking, fast and slow* (Vol. 1). Farrar, Straus and Giroux New York.
Kalashnikova, M., Mattock, K., & Monaghan, P. (2015). The effects of linguistic experience on the flexible use of mutual exclusivity in word learning. *Bilingualism: Language and Cognition*, *18*(4), 626–638.
Kandel, S., Burfin, S., Méary, D., Ruiz-Tada, E., Costa, A., & Pascalis, O. (2016). The impact of early bilingualism on face recognition processes. *Frontiers in psychology*, *7*, 1080.
Kao, J., Bergen, L., & Goodman, N. (2014). Formalizing the pragmatics of metaphor understanding. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 36).
Katerelos, M., Poulin-Dubois, D., & Oshima-Takane, Y. (2011). A cross-linguistic study of word-mapping in 18-to 20-month-old infants. *Infancy*, *16*(5), 508–534.
Katsos, N. (2009). Evaluating under-informative utterances with context-dependent and context-independent scales: experimental and theoretical implications. *Experimental semantics and pragmatics*, 51–73.
Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81.
Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, *16*(4), 705–710.
Kaushanskaya, M., & Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic bulletin & review*, *19*(5), 935–941.
Kersten, A. W., & Smith, L. B. (2002). Attention to novel objects during verb learning. *Child development*, *73*(1), 93–109.
Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in

conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.
King, B. J. (1994). *The information continuum: Evolution of social information transfer in monkeys, apes, and hominids.* School of American Research Press.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review*, *95*(2), 163.
Kissine, M. (2012). Pragmatics, cognitive flexibility and autism spectrum disorders. *Mind & language*, *27*(1), 1–28.
Kissine, M., Cano-Chervel, J., Carlier, S., De Brabanter, P., Ducenne, L., Pairon, M.-C., ... Leybaert, J. (2015). Children with autism understand indirect speech acts: evidence from a semi-structured act-out task. *PLoS One*, *10*(11), e0142191.
Koenig, M., & Woodward, A. L. (2012). Toddlers learn words in a foreign language: The role of native vocabulary knowledge. *Journal of child language*, *39*(2), 322–337.
Kovács, Á. M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*, *106*(16), 6556–6560.
Kovács, Á. M., & Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *science*, *325*(5940), 611–612.
Kronmüller, E., Morisseau, T., & Noveck, I. A. (2014). Show me the pragmatic contribution: a developmental investigation of contrastive inference. *Journal of child language*, *41*(5), 985–1014.
Kuhl, P. K. (2007). Is speech learning 'gated'by the social brain? *Developmental science*, *10*(1), 110–120.
Kurumada, C., & Clark, E. V. (2017). Pragmatic inferences in context: learning to interpret contrastive prosody. *Journal of child language*, *44*(4), 850–880.
Lane, L. W., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1466.
Levinson, S. C. (1987). Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of linguistics*, *23*(2), 379–434.
Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* MIT press.
Lewis, M. L., & Frank, M. C. (2016). Understanding the effect of social context on learning: A replication of xu and tenenbaum (2007b). *Journal of Experimental Psychology: General*, *145*(9), e72.
Lewis, M. L., & Frank, M. C. (2018). Still suspicious: the suspicious-coincidence effect revisited. *Psychological science*, *29*(12), 2039–2047.
Liss, P. H., & Haith, M. M. (1970). The speed of visual processing in children and adults: Effects of backward and forward masking. *Perception & Psychophysics*, *8*(6), 396–398.
Lorge, I., & Katsos, N. (2018). Listener-adapted speech: Bilinguals adapt in a more sensitive way. *Linguistic Approaches to Bilingualism*.
Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628–634.
Marian, V. (2009). Audio-visual integration during bilingual language processing. *The*

bilingual mental lexicon: Interdisciplinary approaches, 52–78.
Marian, V., & Kaushanskaya, M. (2007). Cross-linguistic transfer and borrowing in bilinguals. *Applied Psycholinguistics*, *28*(2), 369–390.
Markman, E. M. (1987). How children constrain the possible meanings of words.
Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive science*, *14*(1), 57–77.
Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua*, *92*, 199–227.
Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive psychology*, *16*(1), 1–27.
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*(3), 241–275.
McKellin, W. H., Shahin, K., Hodgson, M., Jamieson, J., & Pichora-Fuller, K. (2007). Pragmatics of conversation and communication in noisy settings. *Journal of Pragmatics*, *39*(12), 2159–2184.
Menjivar, J., & Akhtar, N. (2017). Language experience and preschoolers' foreign word learning. *Bilingualism: Language and Cognition*, *20*(3), 642–648.
Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the society for research in child development*, i–129.
Mervis, C. B. (1987). Child-basic object categories and early lexical development.
Moll, H., & Tomasello, M. (2007). How 14-and 18-month-olds know what others have experienced. *Developmental psychology*, *43*(2), 309.
Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental psychology*, *35*(1), 60.
Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of experimental child psychology*, *114*(2), 187–202.
Morisseau, T., Davies, C., & Matthews, D. (2013). How do 3-and 5-year-olds respond to under-and over-informative utterances? *Journal of Pragmatics*, *59*, 26–39.
Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child development*, *72*(3), 834–843.
Muhlstein, L. (2016). A minimal parsimonious model of pragmatic comprehension.
Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child development*, *74*(1), 221–237.
Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336.
Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of child language*, *17*(2), 357–374.
Nappa, R., January, D., Gleitman, L., & Trueswell, J. (2004). Paying attention to attention: Perceptual priming effects on word order. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 26).
Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of experimental child psychology*, *110*(1), 52–61.
Noble, C. H., Rowland, C. F., & Pine, J. M. (2011). Comprehension of argument

structure and semantic roles: Evidence from english-learning children and the forced-choice pointing paradigm. *Cognitive science*, *35*(5), 963–982.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

Nunberg, G. (1987). Poetic and prosaic metaphors. *Theoretical Issues in Natural Language Processing 3*.

Nurmsoo, E., & Bloom, P. (2008). Preschoolers' perspective taking in word learning: Do they blindly follow eye gaze? *Psychological Science*, *19*(3), 211–215.

Oshima-Takane, Y., Ariyama, J., Kobayashi, T., Katerelos, M., & Poulin-Dubois, D. (2011). Early verb learning in 20-month-old japanese-speaking children. *Journal of child language*, *38*(3), 455–484.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278.

Paatsch, L. E., & Toe, D. M. (2013). A comparison of pragmatic abilities of children who are deaf or hard of hearing and their hearing peers. *Journal of deaf studies and deaf education*, *19*(1), 1–19.

Palmer, S. B., Fais, L., Golinkoff, R. M., & Werker, J. F. (2012). Perceptual narrowing of linguistic sign occurs in the 1st year of life. *Child development*, *83*(2), 543–553.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253–282.

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition*, *12*(1), 71–82.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology Section A*, *48*(1), 98–107.

Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in second language acquisition*, *18*(1), 1–25.

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied psycholinguistics*, *18*(1), 41–58.

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

Perkins, M. R. (1998). Is pragmatics epiphenomenal? evidence from communication disorders. *Journal of Pragmatics*, *29*(3), 291–311.

Peterson, C. C., & Siegal, M. (1995). Deafness, conversation and theory of mind. *Journal of child Psychology and Psychiatry*, *36*(3), 459–474.

Pierrehumbert, J., & Hirschberg, J. B. (1990). The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication* (pp. 271–311).

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-specific generalization of pragmatic inferences based on under-and over-informative prenominal adjective use. *Frontiers in psychology*, *6*, 2035.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological science*, *26*(4), 490–498.

Poulin-Dubois, D., Graham, S., & Sippola, L. (1995). Early lexical development: The

contribution of parental labelling and infants' categorization abilities. *Journal of Child Language*, *22*(2), 325–343.
Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child development*, *83*(1), 236–250.
Quine, W. V. O. (1960). Word and object (studies in communication). *New York and London: Tech-nology Press of MIT*.
Rach, S., & Diederich, A. (2006). Visual–tactile integration: does stimulus duration influence the relative amount of response enhancement? *Experimental brain research*, *173*(3), 514–520.
Rapoport, A., Chammah, A. M., & Orwant, C. J. (1965). *Prisoner's dilemma: A study in conflict and cooperation* (Vol. 165). University of Michigan press.
Recanati, F. (2004). *Literal meaning*. Cambridge University Press.
Rettenbach, R., Diller, G., & Sireteanu, R. (1999). Do deaf people see better? texture segmentation and visual search compensate in adult but not in juvenile subjects. *Journal of Cognitive Neuroscience*, *11*(5), 560–583.
Ricciardelli, L. A. (1992). Bilingualism and cognitive development in relation to threshold theory. *Journal of psycholinguistic research*, *21*(4), 301–316.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382–439.
Ross, J., & Melinger, A. (2017). Bilingual advantage, bidialectal advantage or neither? comparing performance across three tests of executive function in middle childhood. *Developmental science*, *20*(4), e12405.
Rowe, L., Jacobson, R., & Saylor, M. M. (2015). Differences in how monolingual and bilingual children learn second labels for familiar objects. *Journal of child language*, *42*(6), 1219–1236.
Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? an efficiency-based analysis of color overspecification. *Frontiers in psychology*, *7*, 153.
Rubio-Fernández, P., & Glucksberg, S. (2012). Reasoning about other people's beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 211.
Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of akhtar, carpenter, and tomasello. *Child development*, *69*(1), 94–104.
Scheffner Hammer, C., Lawrence, F. R., & Miccio, A. W. (2008). Exposure to english before and after entry into head start: Bilingual children's receptive language growth in spanish and english. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 30–56.
Schenkman, B. N., & Nilsson, M. E. (2010). Human echolocation: Blind and sighted persons' ability to detect sounds recorded in the presence of a reflecting object. *Perception*, *39*(4), 483–501.
Scofield, J., & Behrend, D. A. (2011). Clarifying the role of joint attention in early word learning. *First Language*, *31*(3), 326–341.
Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science*, *23*(9), 994–999.
Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of psycholinguistic*

research, *32*(1), 3–23.
Sedivy, J. C., Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Carlson, G. (1995). Using intonationally-marked presuppositional information in on-line language processing: Evidence from eye movements to a visual model. In *Proceedings of the 17th annual conference of the cognitive science society* (pp. 375–380).
Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.
Sekerina, I. A., & Trueswell, J. C. (2012). Interactive processing of contrastive expressions by russian children. *First language*, *32*(1-2), 63–87.
Shipley, E. F., & Kuhn, I. F. (1983). A constraint on comparisons: Equally detailed alternatives. *Journal of Experimental Child Psychology*, *35*(2), 195–222.
Shneidman, L. A., Buresh, J. S., Shimpi, P. M., Knight-Schwarz, J., & Woodward, A. L. (2009). Social experience, social attention and word learning in an overhearing paradigm. *Language Learning and Development*, *5*(4), 266–281.
Shneidman, L. A., Gweon, H., Schulz, L. E., & Woodward, A. L. (2016). Learning from others and spontaneous exploration: A cross-cultural investigation. *Child Development*, *87*(3), 723–735.
Shook, A., Goldrick, M., Engstler, C., & Marian, V. (2015). Bilinguals show weaker lexical access during spoken sentence comprehension. *Journal of psycholinguistic research*, *44*(6), 789–802.
Siegal, M., Iozzi, L., & Surian, L. (2009). Bilingualism and conversational understanding in young children. *Cognition*, *110*(1), 115–122.
Siegal, M., Matsuo, A., Pond, C., & Otsu, Y. (2007). Bilingualism and cognitive development: Evidence from scalar implicatures. In *Proceedings of the eighth tokyo conference on psycholinguistics* (pp. 265–280).
Siegal, M., Surian, L., Matsuo, A., Geraci, A., Iozzi, L., Okumura, Y., & Itakura, S. (2010). Bilingualism accentuates children's conversational understanding. *PloS one*, *5*(2), e9004.
Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6–18.
Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.
Song, H.-j., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental psychology*, *44*(6), 1789.
Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592.
Spencer, J. P., Perone, S., Smith, L. B., & Samuelson, L. K. (2011). Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological science*, *22*(8), 1049–1057.
Sperber, D. (1994). Understanding verbal understanding. *What is intelligence*, 179–198.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Cambridge, MA, USA: Harvard University Press.
Starlinger, I., & Niemeyer, W. (1981). Do the blind hear better? investigations on auditory processing in congenital or early acquired blindness i. peripheral functions. *Audiology*, *20*(6), 503–509.
Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool

children. *Language Learning and Development*, *11* (2), 176–190.
Storms, G., Ameel, E., & Malt, B. C. (2015). Development of cross-language lexical influence. *International Journal of Bilingual Education and Bilingualism*, *18* (5), 529–547.
Sullivan, J., Davidson, K., Wade, S., & Barner, D. (2017). Differentiating scalar implicature from exclusion inferences in language acquisition.
Sullivan, M. D., Janus, M., Moreno, S., Astheimer, L., & Bialystok, E. (2014). Early stage second-language learning improves executive control: Evidence from erp. *Brain and language*, *139*, 84–98.
Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, *26* (2), 212–215.
Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological science*, *18* (7), 580–586.
Surian, L., Tedoldi, M., & Siegal, M. (2010). Sensitivity to conversational maxims in deaf and hearing children. *Journal of child language*, *37* (4), 929–943.
Syrett, K., Austin, J., Sanchez, L., Germak, C., Lingwall, A., Perez-Cortes, S., ... Baker, H. (2017). The influence of conversational context and the developing lexicon on the calculation of scalar implicatures. *Linguistic Approaches to Bilingualism*, *7* (2), 230–264.
Syrett, K., Lingwall, A., Perez-Cortes, S., Austin, J., Sánchez, L., Baker, H., ... Arias-Amaya, A. (2017). Differences between spanish monolingual and spanish-english bilingual children in their calculation of entailment-based scalar implicatures. *Glossa: a journal of general linguistics*, *2* (1).
Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? *Scientific Reports*, *4*, 5304.
Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: salience, absoluteness, and consistency. *Frontiers in psychology*, *6*, 1703.
Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, *24* (4), 629–640.
Teresa Guasti, M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and cognitive processes*, *20* (5), 667–696.
Tincoff, R. (2001). Infants' attention to speech and non-speech vocalizations, and its relation to word-learning. In *Poster presented at the biennial meeting of the society for research in child development, minneapolois, mn, april.*
Titone, D. A., & Connine, C. M. (1994). Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. *Metaphor and Symbol*, *9* (4), 247–270.
Titone, D. A., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of pragmatics*, *31* (12), 1655–1674.
Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, *10* (4), 401–413.
Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, *10* (2), 201–224.
Tomasello, M., & Kruger, A. C. (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of child language*, *19* (2), 311–333.

Tomasello, M., Moore, C., & Dunham, P. J. (1995). Joint attention: Its origins and role in development. *Joint Attention as Social Cognition. Hillsdale, NJ: Lawrence Erlbaum Associates*, 103–130.
Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of child language*, *23*(1), 157–176.
Triesch, J., Teuscher, C., Deák, G. O., & Carlson, E. (2006). Gaze following: why (not) learn it? *Developmental science*, *9*(2), 125–147.
Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of experimental psychology: General*, *113*(2), 169.
Umbel, V. M., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child development*, *63*(4), 1012–1020.
Volden, J., & Phillips, L. (2010). Measuring pragmatic language in speakers with autism spectrum disorders: Comparing the children's communication checklist—2 and the test of pragmatic language. *American Journal of Speech-Language Pathology*.
Warmington, M. A., Kandru-Pothineni, S., & Hitch, G. J. (2018). Novel-word learning, executive control and working memory: A bilingual advantage. *Bilingualism: Language and Cognition*.
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., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive psychology*, *43*(3), 217–242.
Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in cognitive sciences*, *13*(6), 258–263.
Waxman, S. R., & Kosowski, T. D. (1990). Nouns mark category relations: Toddlers' and preschoolers' word-learning biases. *Child development*, *61*(5), 1461–1473.
Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*(3), 367–392.
Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*(3), 239–277.
Werker, J. (2012). Perceptual foundations of bilingual acquisition in infancy. *Annals of the New York Academy of Sciences*, *1251*(1), 50–61.
Wermelinger, S., Gampe, A., & Daum, M. M. (2017). Bilingual toddlers have advanced abilities to repair communication failure. *Journal of experimental child psychology*, *155*, 84–94.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.
Woodward, A. (1993). The effect of labeling on children's attention to objects. In *Proceedings of the 24th annual child language research forum* (pp. 35–47).
Xu, F., & Tenenbaum, J. B. (2007a). Sensitivity to sampling in bayesian word learning. *Developmental science*, *10*(3), 288–297.
Xu, F., & Tenenbaum, J. B. (2007b). Word learning as bayesian inference. *Psychological review*, *114*(2), 245.
Yost, W. A. (1997). The cocktail party problem: Forty years later. *Binaural and spatial hearing in real and virtual environments*, 329–347.

Yow, W. Q. (2015). Monolingual and bilingual preschoolers' use of gestures to interpret ambiguous pronouns. *Journal of child language*, *42*(6), 1394–1407.

Yow, W. Q., Li, X., Lam, S., Gliga, T., Chong, Y. S., Kwek, K., & Broekman, B. F. (2017). A bilingual advantage in 54-month-olds' use of referential cues in fast mapping. *Developmental science*, *20*(1), e12482.

Yow, W. Q., & Markman, E. M. (2011a). Bilingualism and children's use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition*, *14*(4), 562–569.

Yow, W. Q., & Markman, E. M. (2011b). Young bilingual children's heightened sensitivity to referential cues. *Journal of Cognition and Development*, *12*(1), 12–31.

Yow, W. Q., & Markman, E. M. (2015). A bilingual advantage in how children integrate multiple cues to understand a speaker's referential intent. *Bilingualism: Language and Cognition*, *18*(3), 391–399.

Yow, W. Q., & Markman, E. M. (2016). Children increase their sensitivity to a speaker's nonlinguistic cues following a communicative breakdown. *Child development*, *87*(2), 385–394.

Zelazo, P. D. (2006). The dimensional change card sort (dccs): A method of assessing executive function in children. *Nature protocols*, *1*(1), 297.

Zinszer, B. D., Rolotti, S. V., Li, F., & Li, P. (2018). Bayesian word learning in multiple language environments. *Cognitive science*, *42*, 439–462.

Zosh, J. M., Brinster, M., & Halberda, J. (2013). Optimal contrast: Competition between two referents improves word learning. *Applied Developmental Science*, *17*(1), 20–28.

Appendix A

Stimuli

A.1 Experiment 1

Puppet's story:

'This is my friend Mr. Puppet. He has made a lot of new alien friends with strange names and he has given them some toys, but now he needs them back. Can you help him get his toys back by finding the right alien?'

Warm-up:

'It's very important that you touch the right alien or Mr. Puppet will never find his toys. You have to be careful to choose the right alien, some aliens might be dirty, or wet. For example, here are some gloops. Can you touch the wet gloop? And the dirty gloop?'

Fig. A.1 Warm-up stimulus Experiment 1

Table A.1 audio stimuli experiment 1

control	display
'Touch the wet flurg'	four aliens (two of the same kind, two different ones),
'Touch the dry dinkoo'	one wet one and three dry ones
'Touch the clean patam'	four aliens (two of the same kind, two different ones),
'Touch the dirty tweep'	one clean one and three dirty ones
stressed modifier	display
'Touch the WET plonk'	four aliens (two of the same kind, two different ones),
'Touch the DRY yubba'	two wet ones and two dry ones
'Touch the CLEAN moozie'	four aliens (two of the same kind, two different ones),
'Touch the DIRTY ral'	two clean ones and two dirty ones
unstressed modifier	display
'Touch the wet gorp'	four aliens (two of the same kind, two different ones),
'Touch the dry pitack'	two wet ones and two dry ones
'Touch the clean rapook'	four aliens (two of the same kind, two different ones),
'Touch the dirty lep'	two clean ones and two dirty ones
(list 2)	display
same audio files	stressed/unstressed pictures switched

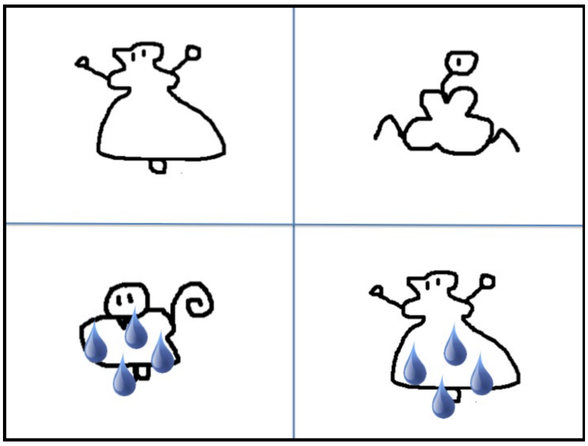


Fig. A.2 stressed/non stressed condition

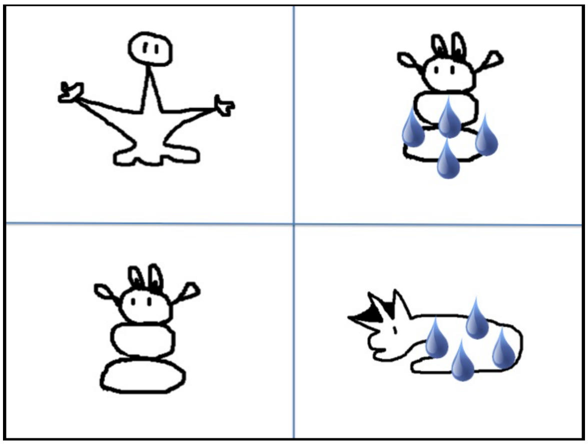


Fig. A.3 stressed/non stressed condition

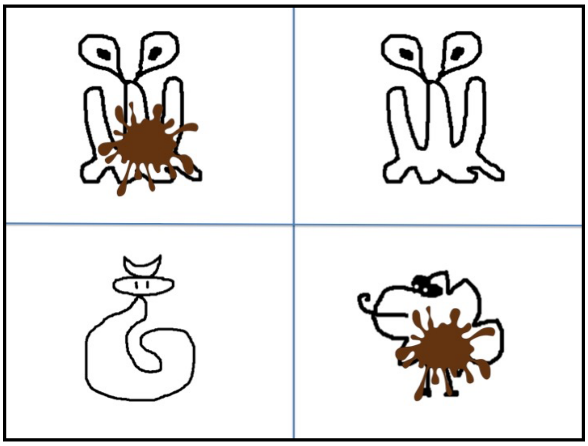
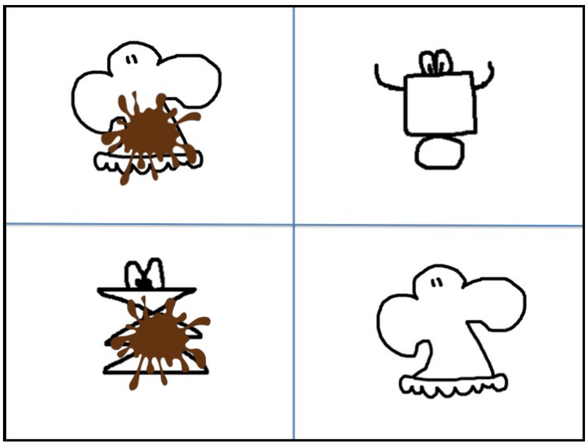


Fig. A.4 stressed/non stressed condition



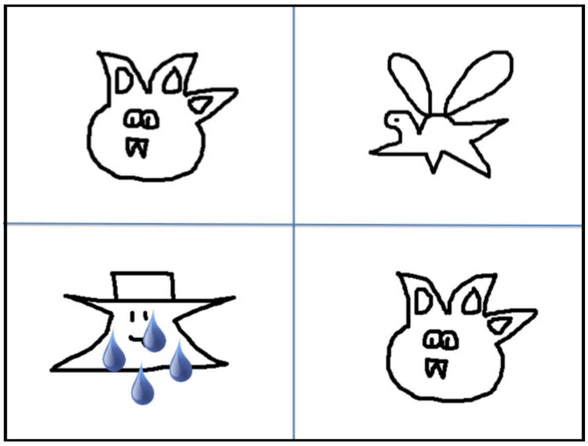


Fig. A.10 control condition

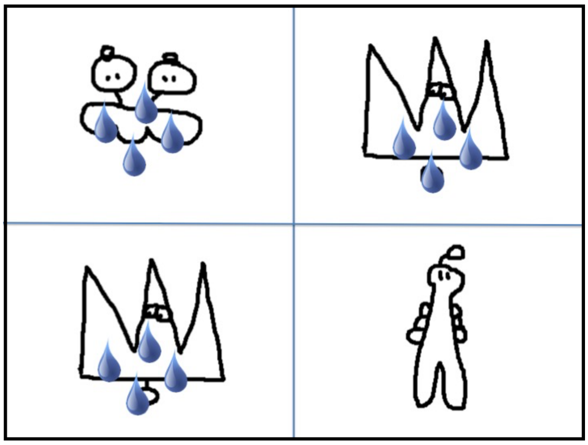


Fig. A.11 control condition

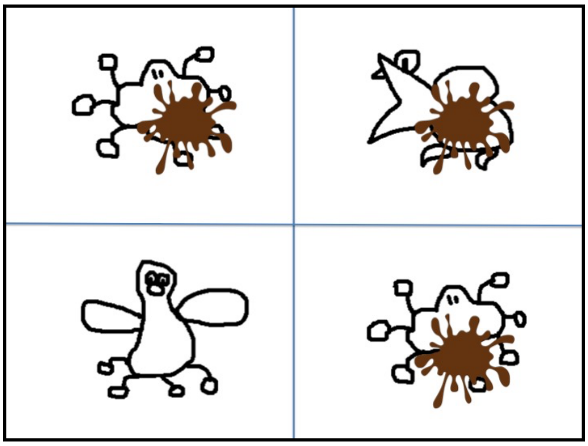
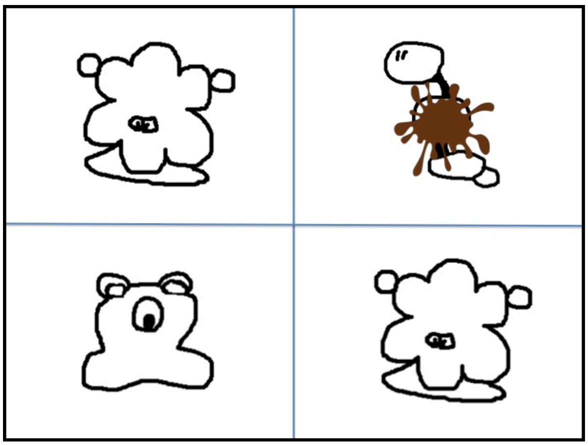


Fig. A.12 control condition



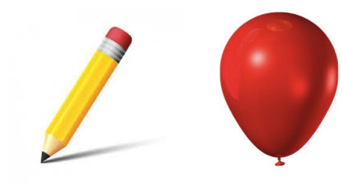


Fig. A.14 Warm-up stimulus Experiment 2 *'Oh, look at these! Have you seen these?'*

Fig. A.15 Warm-up stimulus Experiment 2 *'Can you show me the pencil?'*

Fig. A.16 Warm-up stimulus Experiment 2 *'Can you show me the spotik?'*

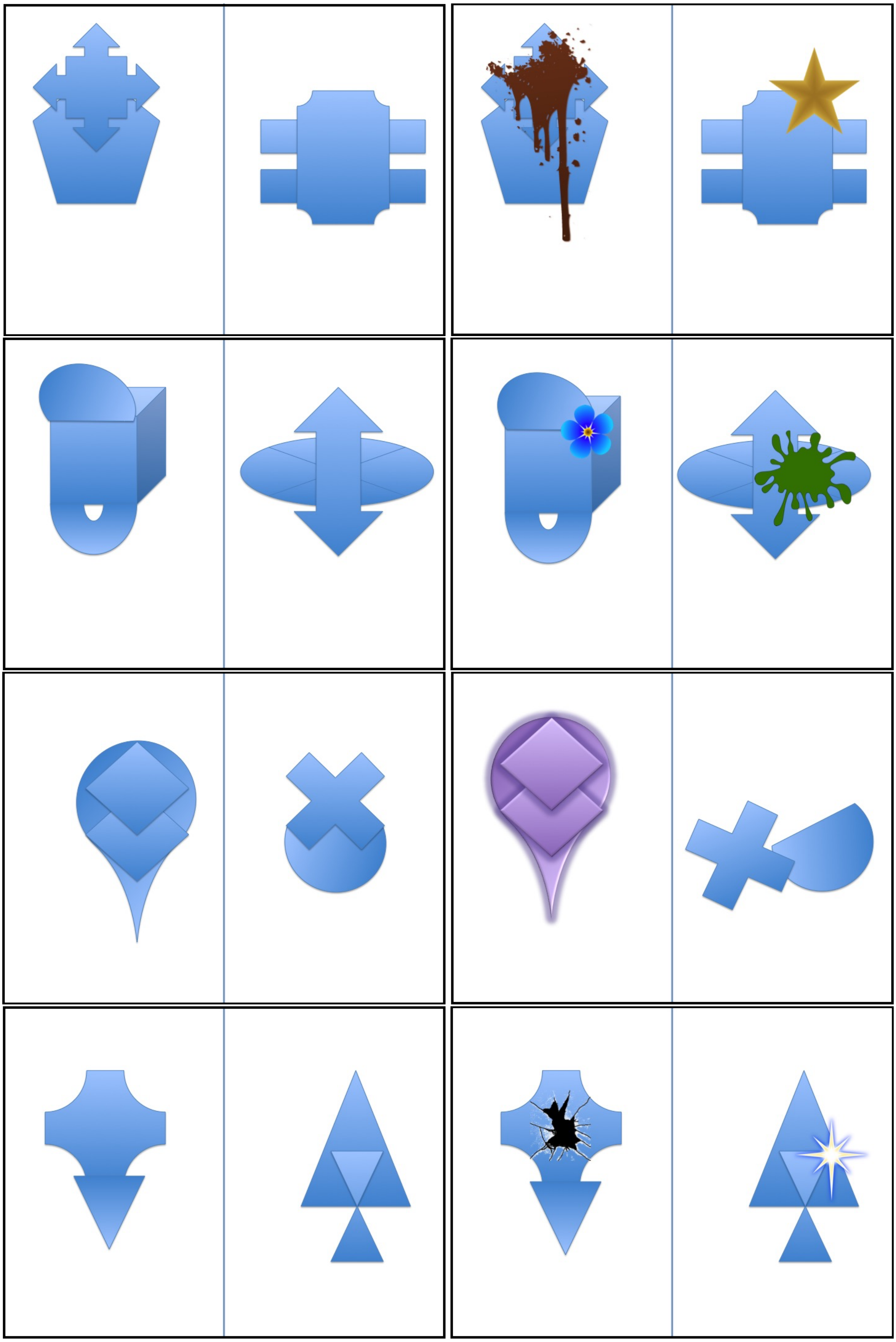
A.2 Experiment 2

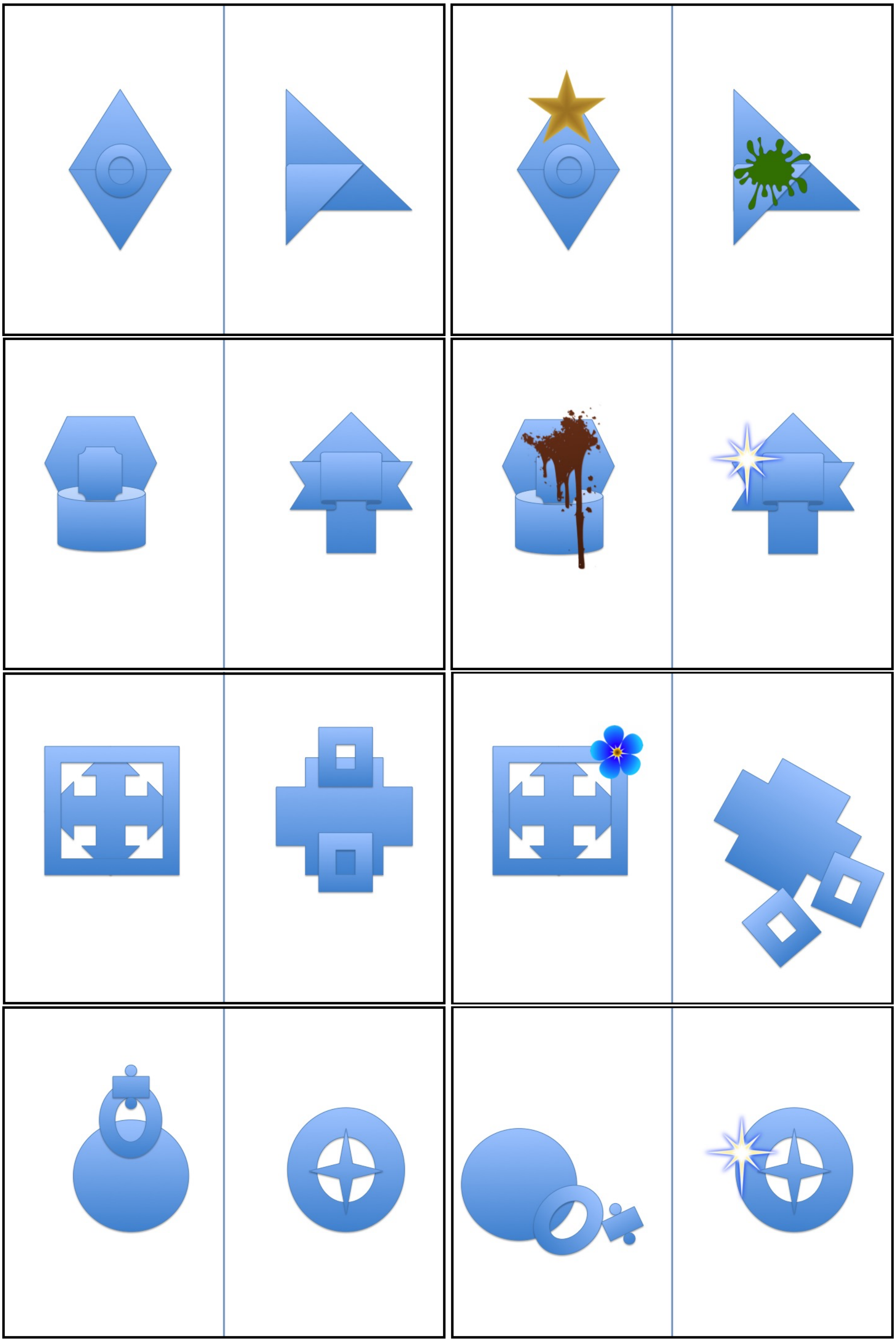
Puppet's story:

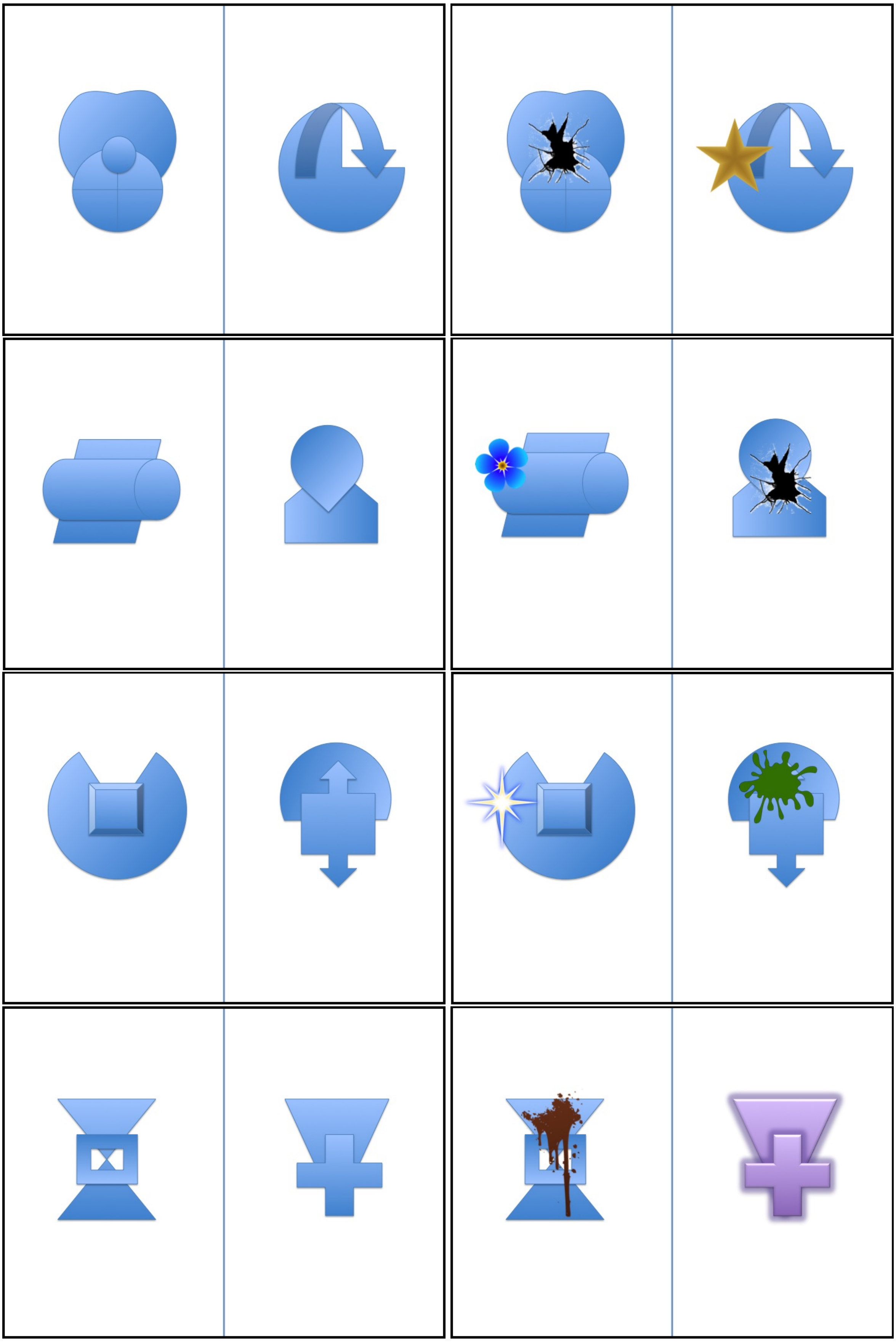
'Do you want to see more objects that the aliens brought? Oh, but Mr. Puppet has been playing with them, so some of them got dirty or broken, but he also decorated some with pretty things'

Table A.2 audio stimuli experiment 2

preparation	**neutral affect**	display
'Oh look at these have you seen these'	(neutral tone) *'Oh, look at the spoodle, can you touch the spoodle?'*	mud/star
'Oh look at these have you seen these'	(neutral tone) *'Oh, look at the nurmy, can you touch the nurmy?'*	flower/slime
'Oh look at these have you seen these'	(neutral tone) *'Oh, look at the goti, can you touch the goti?'*	pink/broken
'Oh look at these have you seen these'	(neutral tone) *'Oh, look at the fopal, can you touch the fopal?'*	hole/spark
preparation	**positive affect**	
'Oh look at these have you seen these'	(happy tone) *'Oh, look at the figoo, can you touch the figoo?'*	star/slime
'Oh look at these have you seen these'	(happy tone) *'Oh, look at the dazee, can you touch the dazee?'*	mud/spark
'Oh look at these have you seen these'	(happy tone) *'Oh, look at the grof, can you touch the grof?'*	flower/broken
'Oh look at these have you seen these'	(happy tone) *'Oh, look at the pilk, can you touch the pilk?'*	broken/spark
preparation	**negative affect**	
'Oh look at these have you seen these'	(sad tone) *'Oh, look at the zarp, can you touch the zarp?'*	hole/star
'Oh look at these have you seen these'	(sad tone) *'Oh, look at the mook, can you touch the mook?'*	flower/hole
'Oh look at these have you seen these'	(sad tone) *'Oh, look at the slarp, can you touch the slarp?'*	spark/slime
'Oh look at these have you seen these'	(sad tone) *'Oh, look at the klem, can you touch the klem?'*	mud/pink
preparation	**list 2**	
same audio files	pictures switched between conditions	







A.3 Experiment 3

Puppet's story:

'Let's carry on helping Mr. Puppet! The aliens left a bag with objects from their planet and animals have stolen some of them. Can you help Mr. Puppet find the animals that have taken the aliens' objects?'

A.4 Experiment 5

Puppet's story (teacher condition) *'Mr. Puppet wants to show you some of his new toys'*/(learner condition) *'I found the bag that the aliens left! Let's shake it and see what funny objects are inside'*

A.5 Experiment 6

Puppet's story: *Now we're going to watch Mr. Puppet's friend Katie play with some of his toys'*

Table A.3 audio stimuli experiment 3a

control	display
'Oh, a kitten that has a cup!	two kittens, one with cup, one with novel object
Can you touch the FROG that has a cup?'	two frogs, one with cup, one with novel object
'Oh, a dog that has a clock!	two dogs, one with clock, one with novel object
Can you touch the FROG that has a clock?'	two frogs, one with clock, one with novel object
'Oh, a bird that has a tree!	two birds, one with tree, one with novel object
Can you touch the FROG that has a tree?'	two frogs, one with tree, one with novel object
'Oh, a fish that has a chair!	two birds, one with chair, one with novel object
Can you touch the FROG that has a chair?'	two frogs, one with chair, one with novel object

disambiguation	display
'Oh, a kitten that has a gabo!	two kittens, one with pencil, one with novel object
Can you touch the FROG that has a gabo?'	two frogs, one with pencil, one with novel object
'Oh, a dog that has a keef	two dogs, one with apple, one with novel object
Can you touch the FROG that has a keef?'	two frogs, one with apple, one with novel object
'Oh, a bird that has a fid!	two birds, one with carrot, one with novel object
Can you touch the FROG that has a fid?'	two frogs, one with carrot, one with novel object
'Oh, a fish that has a razee!	two birds, one with shoe, one with novel object
Can you touch the FROG that has a razee?'	two frogs, one with shoe, one with novel object

frequency	display
'Oh, a kitten that has a fep!	two kittens, one with object 1, one with objects 1, 2
Can you touch the FROG that has a fep?'	two frogs, one with object 1, one with object 2
'Oh, a dog that has a toma	two dogs, one with object 1, one with objects 1, 2
Can you touch the FROG that has a toma?'	two frogs, one with object 1, one with object 2
'Oh, a bird that has a tuki!	two birds, one with object 1, one with objects 1, 2
Can you touch the FROG that has a tuki?'	two frogs, one with object 1, one with object 2
'Oh, a fish that has a zef!	two birds, one with object 1, one with objects 1, 2
Can you touch the FROG that has a zef?'	two frogs, one with object 1, one with object 2

(list 2)	display
same audio files	switched

Table A.4 audio stimuli experiment 3b

control	display
'Oh, a kitten that has a cup!	two kittens, one with cup, one with novel object
Can you touch the FROG that has a cup?'	two frogs, one with cup, one with novel object
'Oh, a dog that has a clock!	two dogs, one with clock, one with novel object
Can you touch the FROG that has a clock?'	two frogs, one with clock, one with novel object
'Oh, a bird that has a tree!	two birds, one with tree, one with novel object
Can you touch the FROG that has a tree?'	two frogs, one with tree, one with novel object
'Oh, a fish that has a chair!	two birds, one with chair, one with novel object
Can you touch the FROG that has a chair?'	two frogs, one with chair, one with novel object

disambiguation	display
'Oh, a kitten that has a gabo!	two kittens, one with pencil, one with novel object
Can you touch the FROG that has a gabo?'	two frogs, one with pencil, one with novel object
'Oh, a dog that has a keef	two dogs, one with apple, one with novel object
Can you touch the FROG that has a keef?'	two frogs, one with apple, one with novel object
'Oh, a bird that has a fid!	two birds, one with carrot, one with novel object
Can you touch the FROG that has a fid?'	two frogs, one with carrot, one with novel object
'Oh, a fish that has a razee!	two birds, one with shoe, one with novel object
Can you touch the FROG that has a razee?'	two frogs, one with shoe, one with novel object

frequency	display
'Oh, the kitten circled in red has a fep!	two kittens, one with object 1, one with objects 1, 2
Can you touch the FROG that has a fep?'	two frogs, one with object 1, one with object 2
'Oh, the dog circled in red has a toma	two dogs, one with object 1, one with objects 1, 2
Can you touch the FROG that has a toma?'	two frogs, one with object 1, one with object 2
'Oh, the bird circled in red has a tuki!	two birds, one with object 1, one with objects 1, 2
Can you touch the FROG that has a tuki?'	two frogs, one with object 1, one with object 2
'Oh, the fish circled in red has a zef!	two birds, one with object 1, one with objects 1, 2
Can you touch the FROG that has a zef?'	two frogs, one with object 1, one with object 2

(list 2)	display
same audio files	switched

Fig. A.17 frequency inference

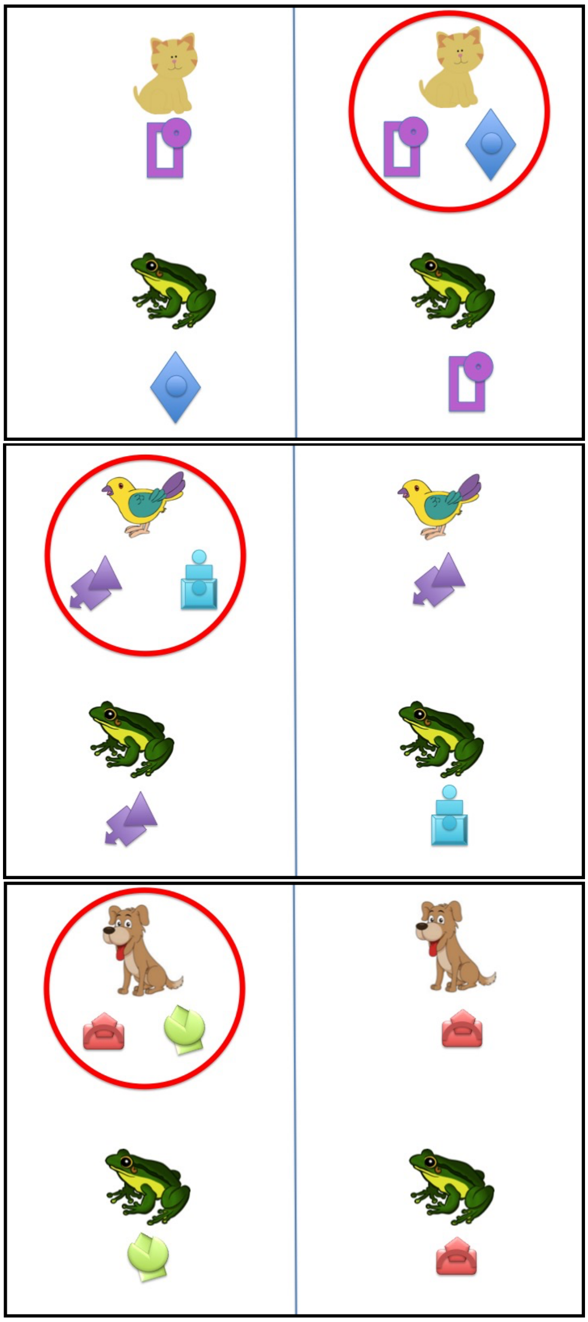


Fig. A.18 mutual exclusivity

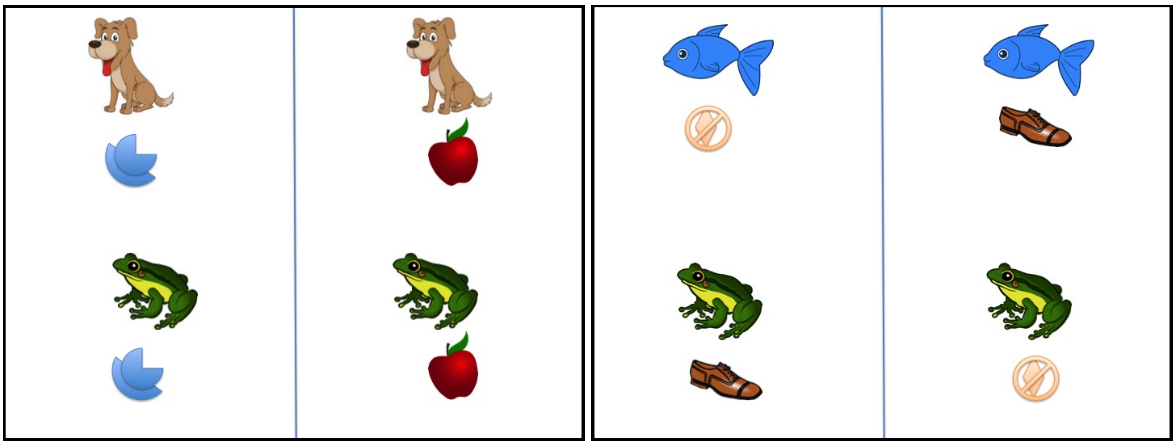


Table A.5 audio stimuli experiment 5a

teacher	display
train: *'Mr. Puppet wants to show you a pok. Look, this is a pok.*	one object
And, look, Mr. Puppet has another pok that he wants to show you!'	one object
test: *'Can you touch the other poks?'*	six objects
train: *'Mr. Puppet wants to show you a skol. Look, this is a skol.*	one object
And, look, Mr. Puppet has another skol that he wants to show you!'	one object
test: *'Can you touch the other skols?'*	six objects
train: *'Mr. Puppet wants to show you a liput. Look, this is a liput.*	one object
And, look, Mr. Puppet has another liput that he wants to show you!'	one object
test: *'Can you touch the other liputs?'*	six objects
train: *'Mr. Puppet wants to show you a murbil. Look, this is a murbil*	one object
And, look, Mr. Puppet has another murbil that he wants to show you!'	one object
test: *'Can you touch the other murbils?'*	six objects

random	display
train: *'Oh, look what fell out of the bag, it's a pok! A pok fell out of the bag!*	one object
Ah, look, another pok fell out of the bag!'	one object
test: *'Can you touch the other poks?'*	six objects
train: *'Oh, look what fell out of the bag, it's a skol! A skol fell out of the bag!*	one object
Ah, look, another skol fell out of the bag!'	one object
test: *'Can you touch the other skols?'*	six objects
train: *'Oh, look what fell out of the bag, it's a liput! A liput fell out of the bag!*	one object
Ah, look, another liput fell out of the bag!'	one object
test: *'Can you touch the other liputs?'*	six objects
train: *'Oh, look what fell out of the bag, it's a murbil! A murbil fell out of the bag!*	one object
Ah, look, another murbil fell out of the bag!'	one object
test: *'Can you touch the other murbils?'*	six objects

(list 2)	display
same audio files	switched

Table A.6 audio stimuli experiment 5b

teacher	display
train: *'Mr. Puppet wants to show you a pok. Look, this is a pok.*	two objects
And, look, Mr. Puppet has another pok that he wants to show you!'	two objects
test: *'Can you touch the poks?'*	six objects
train: *'Mr. Puppet wants to show you a skol. Look, this is a skol.*	two objects
And, look, Mr. Puppet has another skol that he wants to show you!'	two objects
test: *'Can you touch the skols?'*	six objects
train: *'Mr. Puppet wants to show you a liput. Look, this is a liput.*	two objects
And, look, Mr. Puppet has another liput that he wants to show you!'	two objects
test: *'Can you touch the liputs?'*	six objects
train: *'Mr. Puppet wants to show you a murbil. Look, this is a murbil*	two objects
And, look, Mr. Puppet has another murbil that he wants to show you!'	two objects
test: *'Can you touch the murbils?'*	six objects

learner	display
train: *'Oh, look what fell out of the bag, it's a pok! A pok fell out of the bag!*	one object
Can you show me one more pok?'	three objects
test: *'Can you touch the poks?'*	six objects
train: *'Oh, look what fell out of the bag, it's a skol! A skol fell out of the bag!*	one object
Can you show me one more skol?'	three objects
test: *'Can you touch the skols?'*	six objects
train: *'Oh, look what fell out of the bag, it's a liput! A liput fell out of the bag!*	one object
Can you show me one more liput?'	three objects
test: *'Can you touch the liputs?'*	six objects
train: *'Oh, look what fell out of the bag, it's a murbil! A murbil fell out of the bag!*	one object
Can you show me one more murbil?'	three objects
test: *'Can you touch the murbils?'*	six objects

(list 2)	display
same audio files	switched

Fig. A.19 teacher condition

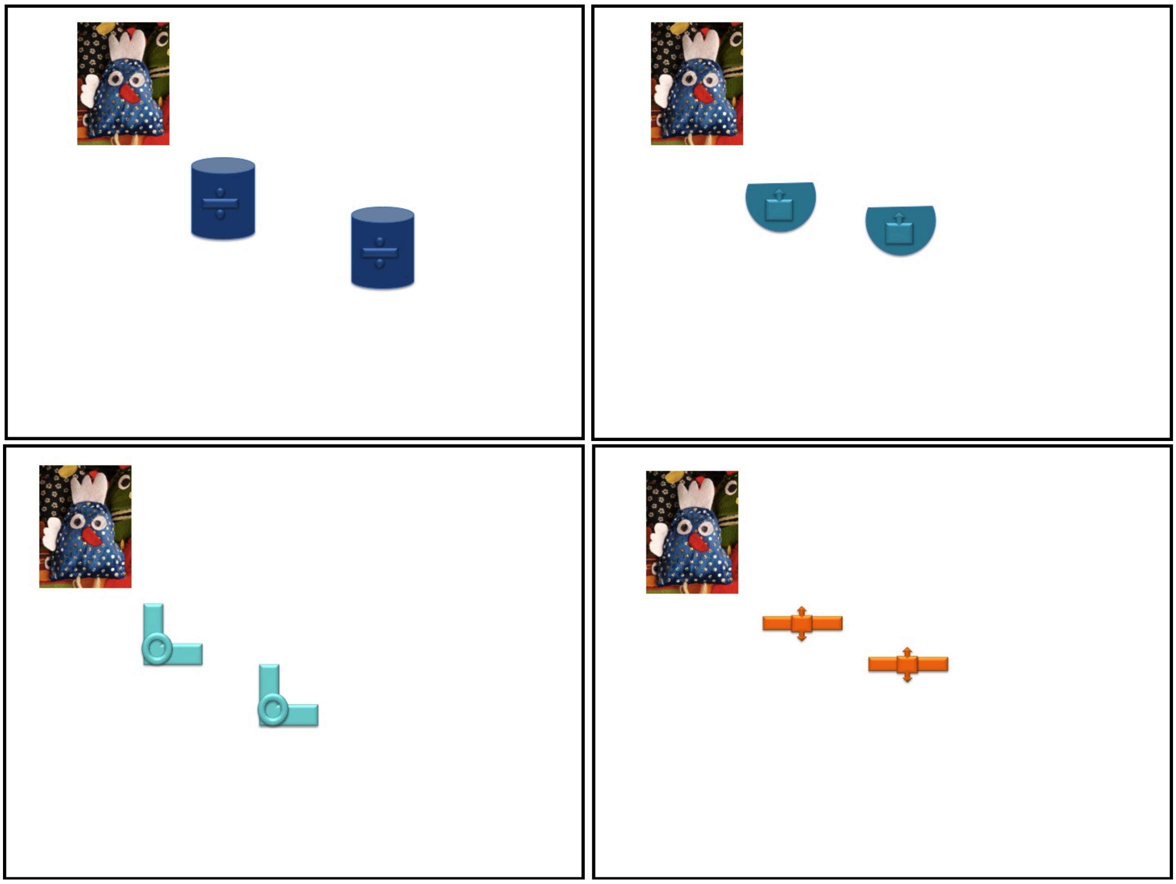


Fig. A.20 learner condition

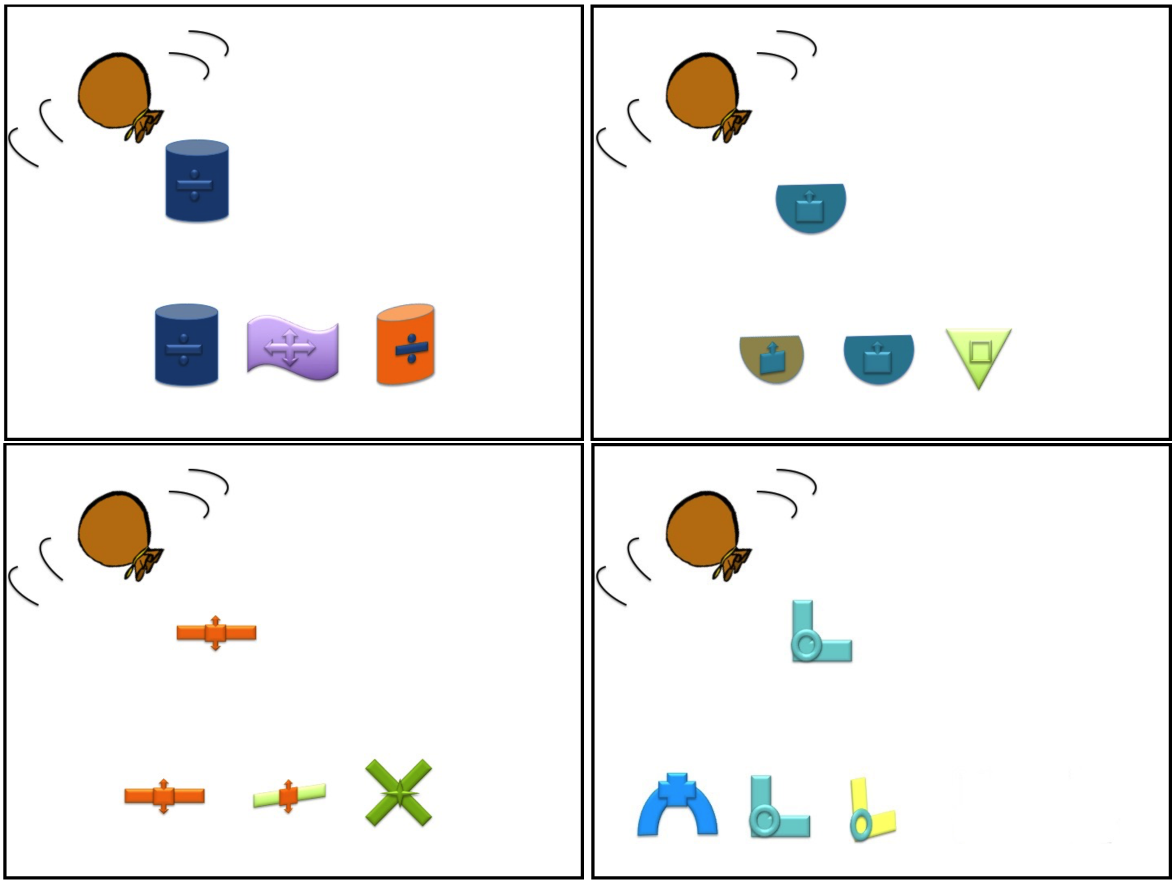


Fig. A.21 random condition

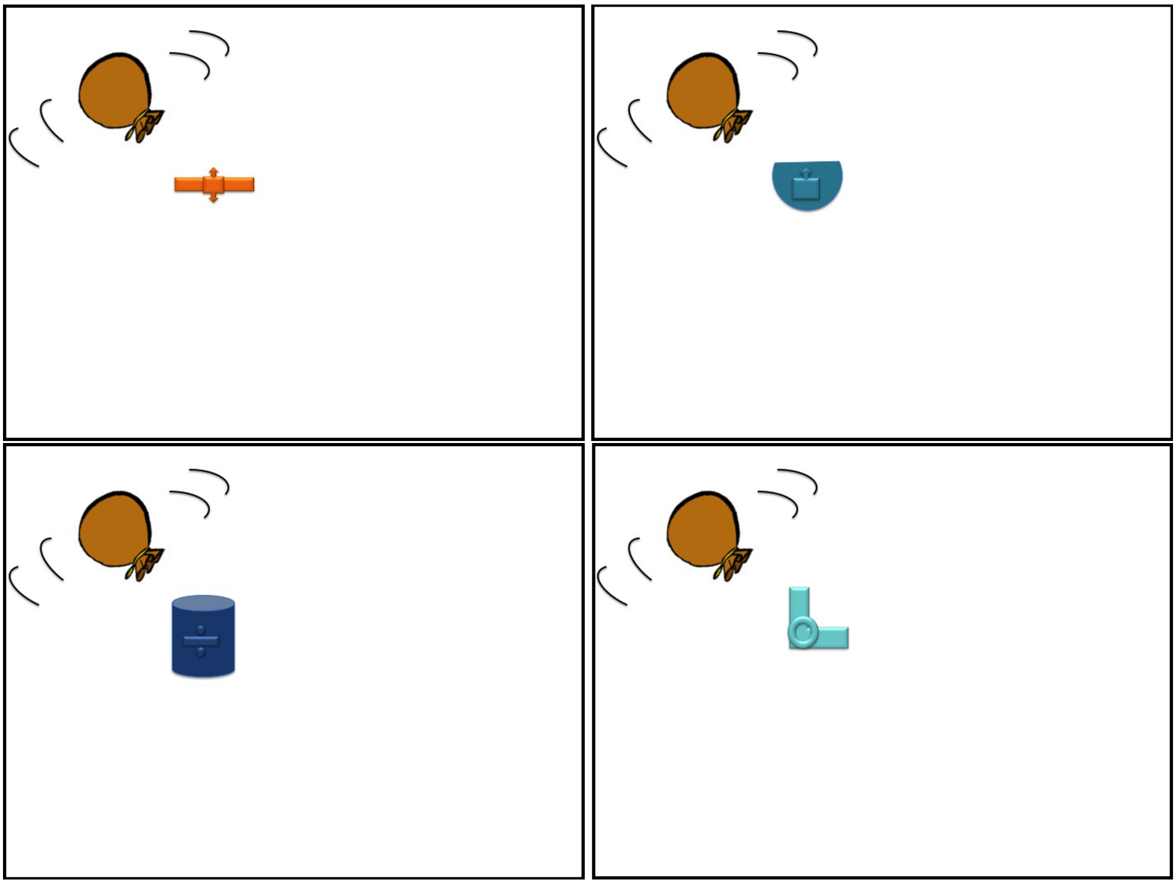


Fig. A.22 test

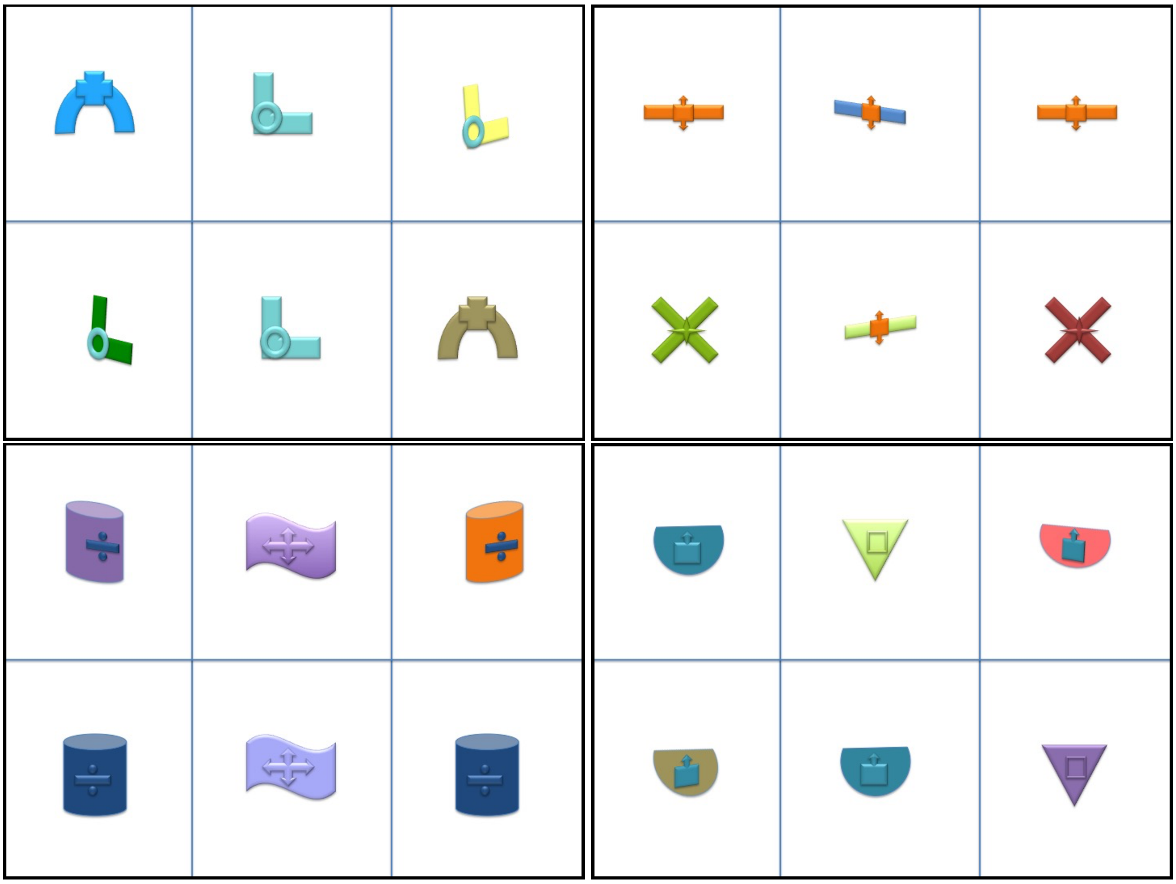


Table A.7 audio stimuli experiment 6a

action highlighted	display
train: *silence*	non-target action with target object
'Dalp'	target action with target object
test: *'Can you show me dalp?'*	target action with other object/other action with target object
train: *silence*	non-target action with target object
'Drook	target action with target object
test: *'Can you show me drook?'*	target action with other object/other action with target object
train: *silence*	non-target action with target object
'Pef'	target action with target object
test: *'Can you show me pef?'*	target action with other object/other action with target object
train: *silence*	non-target action with target object
'Wige'	target action with target object
test: *'Can you show me wige?'*	target action with other object/other action with target object
action neutral	display
train: *'Fup'*	target action with target object
test: *'Can you show me fup?'*	target action with other object/other action with target object
train: *'Dorp*	target action with target object
test: *'Can you show me dorp?'*	target action with other object/other action with target object
train: *'Glay'*	target action with target object
test: *'Can you show me glay?'*	target action with other object/other action with target object
train: *'Squel'*	target action with target object
test: *'Can you show me squel?'*	target action with other object/other action with target object
(list 2)	display
same audio files	switched

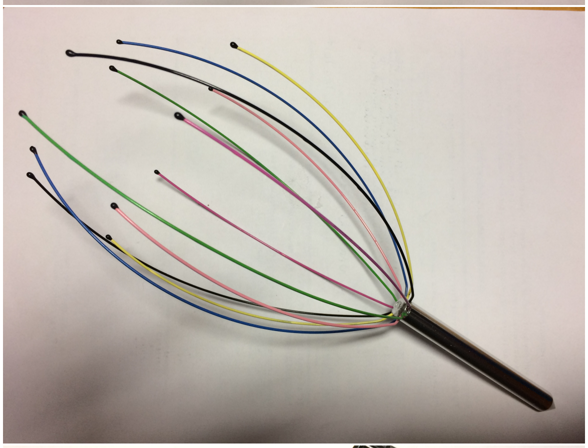
Table A.8 audio stimuli experiment 6b

action highlighted	display
train: *silence*	non-target action with target object
'Dalp'	target action with target object
test: *'Can you show me dalp?'*	target action with other object/other action with target object
train: *silence*	non-target action with target object
'Drook	target action with target object
test: *'Can you show me drook?'*	target action with other object/other action with target object
train: *silence*	non-target action with target object
'Pef'	target action with target object
test: *'Can you show me pef?'*	target action with other object/other action with target object
train: *silence*	non-target action with target object
'Wige'	target action with target object
test: *'Can you show me wige?'*	target action with other object/other action with target object

action neutral	display
train: *'Dalp'*	target action with target object
test: *'Can you show me dalp?'*	target action with other object/other action with target object
train: *'Drook*	target action with target object
test: *'Can you show me drook?'*	target action with other object/other action with target object
train: *'Pef'*	target action with target object
test: *'Can you show me pef?'*	target action with other object/other action with target object
train: *'Wige'*	target action with target object
test: *'Can you show me wige?'*	target action with other object/other action with target object

(list 2)	display
same audio files	switched


BOTANICS





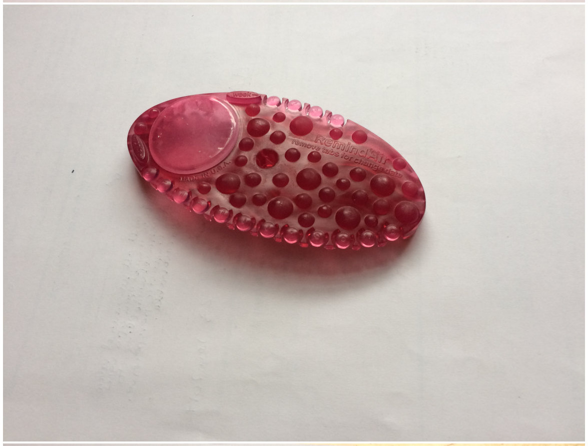



PHILIPS
Satinelle









Appendix B

Parental questionnaire

To be filled in by one of the child's main caregivers (e.g. mother, father, grandparent, guardian, foster carer).

Your relationship to the child (e.g. mother, father, grandmother, guardian, foster carer, etc.):	
Your full name:	
Today's date:	

Note: The data will be anonymised. No one apart from the researchers will have access to your or your child's name. No real names will be used in any report or publication.

Thank you very much for taking the time to fill in this questionnaire!

Please return it to your child's school. If you have any questions, please do not hesitate to be in contact with the researcher, Dr Özge Öztürk - ioo21@cam.ac.uk or +44 (0)1223 760354.

1. General information about the child

Child's full name:	
Child's date of birth:	
Child's age:	
Child's gender:	
Child's place of birth (city/town and country):	

1.6. What language(s) can your child speak and/or understand? Tick for English, and write any additional language(s) in the appropriate box of the table. Please write every language the child can speak/understand even if she/he has little ability in it.

English :	
Other language A :	
Other language B :	
Other language C :	

If you have ticked English only, please proceed to Section 4, 'Information about the Family' (p.6). If you have written down additional languages (A, B, C), please continue with Section 2.

2. Information about the child's language abilities and exposure

2.1. At what age did your child start receiving *consistent and significant* exposure to each of his/her languages?

Language :	**Age in years and months:**
English :	
Other language A :	
Other language B :	
Other language C :	

What we mean by consistent and significant: the child started attending a school (e.g., primary school/kindergarten/daycare) where instruction was held in that language, or one of the main caregivers (e.g., babysitter, caregivers) of the child started to consistently use that language with the child. 2.2. How much is your child exposed to each language? (Please circle one of

the numbers from 1 to 5)

1	2	3	4	5
English never Language(s) A, B and/or C always	English rarely Language(s) A, B and/or C usually	English 50% Language(s) A, B and/or C 50%	English usually Language(s) A, B and/or C rarely	English always Language(s) A, B and/or C never

2.3. How fluent would you say your child is in *speaking* his/her languages? Please tick the appropriate box for each language.

	Not fluent	**Little fluency**	**Somewhat fluent**	**Quite fluent**	**Very fluent**
	Little speaking ability	Some ability; can say short, simple sentences, e.g. answer the phone, or greet a neighbour	Good speaking skills; can express him/herself on many topics, e.g. explain what he/she wants; give instructions	Can use the language adequately in most situations	Very comfortable expressing him/herself in the language in all situations
English					
Language A					
Language B					
Language C					

2.4. How competent would you say your child is in understanding his/her languages for his/her age? Please tick the appropriate box.

	Not competent	**Limited competence**	**Somewhat competent**	**Quite competent**	**Very competent**
	Little understanding ability	Some understanding, e.g. can understand simple conversations	Good understanding of around 50% of communication	Can understand the language adequately for most situations, e.g. follow films or TV shows	Understands almost everything
English					
Language A					
Language B					
Language C					

2.5. How competent would you say your child is in writing in his/her languages? Tick the appropriate box.

	Not competent	**Limited competence**	**Somewhat competent**	**Quite competent**	**Very competent**
	Knows only some of the letters/ characters, e.g. can write only a couple of words	Knows the letters/ characters, e.g. can write their name, some words, and simple combinations of words	Can write simple phrases and sentences but with noticeable mistakes	Can write long sentences, short descriptions with some mistakes	Can have longer written correspondence with friends and family with minor mistakes
English					
Language A					
Language B					
Language C					

2.6. How competent would you say your child is in reading in his/her languages? Tick the appropriate box.

	Not competent	Limited competence	Somewhat competent	Quite competent	Very competent
	Can name some letters/ characters	Can read most or all of the letters/ characters, e.g. a simple shopping list	Can read short pieces of text, e.g. brief messages, notes	Can read longer structured texts, e.g. instructions of board games and video games	Can read complex sentences, paragraphs, short stories
English					
Language A					
Language B					
Language C					

2.7. Does your child codeswitch when they speak? (Please circle) YES / NO (*codeswitching is using English and Language A/B/C words in the same sentence, and is a normal aspect of being multilingual.*)

2.8. Is your child exposed to any of their languages in any of the following ways? (Please circle any that apply)

English	At a language school/club (e.g. Saturday complementary school)	Through meet-ups with friends who speak language A/B/C	Through regular organised activities (e.g. religious meetings, sports clubs)	Through regular contact with relatives
Language A	At a language school/club (e.g. Saturday complementary school)	Through meet-ups with friends who speak language A/B/C	Through regular organised activities (e.g. religious meetings, sports clubs)	Through regular contact with relatives
Language B	At a language school/club (e.g. Saturday complementary school)	Through meet-ups with friends who speak language A/B/C	Through regular organised activities (e.g. religious meetings, sports clubs)	Through regular contact with relatives
Language C	At a language school/club (e.g. Saturday complementary school)	Through meet-ups with friends who speak language A/B/C	Through regular organised activities (e.g. religious meetings, sports clubs)	Through regular contact with relatives

3. Information about ideas about languages

3.1 I want to encourage my child to learn their home language(s)… (Please tick any that apply for languages A, B and/or C)

	Disagree completely	Disagree a bit	Neither agree nor disagree	Agree a bit	Agree completely	Not applicable
...so that they can communicate with family members abroad						
...so that they have better job prospects						
... so that they are part of a home language and culture						
...because it forms a part of their/our religion						
...because it would be a wasted opportunity not to encourage the home language						
... because it introduces them to other cultures						
...because I cannot imagine my children not speaking the home language						

3.2. How proficient are you aiming for your child to be in their home languages? (*While we may all aim at fully equal status of both (or all) languages, there are many different types of multilingualism. Please read the statements below and tick the most appropriate box for each skill.*)

	I don't mind; I will be led by my child in this	I am aiming for this	I really want my child to do this
Basic verbal (listening/speaking) communication			
Being able to hold a good conversation, including more complex topics			
A good amount of verbal vocabulary			
A solid understanding of grammar rules in speaking			
Being able to read basics (books for children a few years below their actual age)			
Advanced reading (books at their actual reading level, or no more than two years below)			
Basic writing (writing short sentences or texts, with errors in spelling)			
Advanced writing (composing own texts, almost secure spelling, one or two years below English level)			
Fully equal status of both or all languages			

3.3. Is your child interested in learning their home language(s) at the moment? YES / NO

3.4. Please indicate how often (if at all) you use the following materials to support home language learning in your family (for languages A, B and/or C).

	daily or nearly daily	a couple of times a week	about once a week	a couple of times a month	about once a month	once every few months	once or twice a year	never used, but heard of it	I am not aware of this
Magazines or newspapers									
Books									
Card or board games (e.g. Top Trumps)									
Live television (e.g. via satellite or cable)									
Subscription television (e.g. Amazon/Netflix)									
Internet-based television (e.g. YouTube) Audio-books									
Websites on the internet									
Language-learning websites									
Internet-based games									
Social Media (e.g. Facebook, MySpace)									
Internet-based communication (e.g. Skype)									
Mobile apps and games for language learning (e.g. MindSnacks, Duolingo)									
Mobile apps and games in the home language (but *not* aimed at language learners)									

4. Information about the family

4.1. Please circle the level of education the child's MOTHER or PRIMARY CAREGIVER has completed

Level	**Please circle**	**Number of years**
Primary school	Yes/No	
Secondary school	Yes/No	
Universty-Bachelor's	Yes/No	
University-Master's	Yes/No	
University-PhD	Yes/No	
Other qualifications	Yes/No	

4.2. What language(s) do YOU speak? (Please write any other languages)

	How did you learn that language? (e.g. from parents in the home, at school, through moving to a new country)	How regularly do you speak or use the language? (never – rarely – sometimes – very often – every day)
English:		
Other:		
Other:		
Other:		

4.3. What language(s) does the child's OTHER MAIN CAREGIVER (if applicable) speak? (Please write any other languages)

	How did she/he learn that language? (e.g. from parents in the home, at school, through moving to a new country)	How regularly does she/he speak or use the language? (never – rarely – sometimes – very often – every day)
English:		
Other:		
Other:		
Other:		

5. Information about school experience

5.1. What does your child like most about being at his/her school?

5.2. What does your child like least about being at his/her school?

5.3. Why did you choose SCHOOL for your child?

5.4. What do you think about the education at your child's school overall?

5.5. I want my child to learn a foreign language at school (please circle): a) as early as possible b) later primary school only c) only in secondary school

5.6. How often do you. . . ? (Please tick)

	Never	Rarely	Sometimes	Very often	Every day
Attend special events at the school					
Talk to child's teacher or headteacher					
Help out in school (e.g., visiting a class, going on a school trip, reading)					
Attend a PTFA meeting					

5.8. What would make the biggest difference in your decision to participate in these activities? (Tick as many as applicable)

If my work schedule allowed it	
If I had transport	
If I had childcare for younger children	
If I felt more comfortable at school	
If there were more support in my home language(s)	
If school staff invited me more	

5.9. How often do you and your child do the following things together in a typical school week?

	Never	Rarely	Sometimes	Very often	Every day
Working on numeracy					
Working on reading/writing in English					
Working on creative activities (e.g. drawing, singing)					
Reading together in English					
Storytelling					

6. Information about your views on language learning

6.1 Thinking of your child and what you know about schooling in England in general, how far do you agree with the following statements? (Please tick)

	Disagree completely	Disagree a bit	Neither agree nor disagree	Agree a bit	Agree completely	Not applicable
I think there are more important things to learn at school than languages						
I think learning another language is an important part of education						
I wish there were more chances for children to learn languages						
Children are spending too much time learning languages						

6.2 I want to encourage my child to learn another language at school... (Please tick)

	Disagree completely	Disagree a bit	Neither agree nor disagree	Agree a bit	Agree completely	Not applicable
...so that they can communicate with people abroad						
...so that they have better job prospects						
...because it introduces them to other cultures						
...because it's an important part of general education						
Other (please specify)						

6.3. How proficient would you like your child to be in another language? (Tick one box for each skill)

	I don't mind; I will be led by my child in this	I am aiming for this	I really want my child to do this
Basic verbal (listening/speaking) communication			
Being able to hold a good conversation, including more complex topics			
A good amount of verbal vocabulary			
A solid understanding of grammar rules in speaking			
Being able to read basics (books for children a few years below their actual age)			
Advanced reading (books at their actual reading level, or no more than two years below)			
Basic writing (writing short sentences or texts, with errors in spelling)			
Advanced writing (composing own texts, almost secure spelling, one or two years below English level)			
Fully equal status of both or all languages			

Thank you very much for completing this questionnaire!