Does multilingualism confer an advantage for pragmatic abilities?

Kyriakos Antoniou[a], Kleanthes Grohmann[b], Maria Kambanaros[c] & Napoleon Katsos[a, 1]

[a]Department of Theoretical and Applied Linguistics, University of Cambridge.
[b]Department of English Studies, University of Cyprus & Cyprus Acquisition Team.
[c]Cyprus Acquisition Team.

[1] Part of this research has been funded by a Euro-XPrag Collaborative Research in Experimental Pragmatics travel grant and an Alexander Onassis Foundation scholarship to the first author.

Abstract

Experimental evidence suggests that bilingualism leads to an advantage in children's executive control (EC), Theory of Mind (ToM) and pragmatic abilities (e.g. Bialystok 2009; Goetz 2003; Siegal et al. 2009; 2010). In this study we aimed to investigate (1) whether multilingualism confers an advantage in children's ability to understand implicatures and (2) whether a potential multilingual advantage in this ability is mediated by some aspect of EC or ToM. In order to achieve these aims we tested multilingual and bilectal children growing up in the Republic of Cyprus where the linguistic situation is typically described as one of diglossia. Children were administered (a) an extensive battery of tasks that cover all aspects of EC (cognitive flexibility, working memory and inhibition) and ToM and (b) a novel and extensive conversational test examining their ability to understand a wide range of implicatures based on Grice's maxims of conversation. Overall, there was only suggestive evidence for a multilingual advantage in pragmatic language. Our results also revealed a multilingual advantage in inhibitory control skills. Finally, we found no evidence for a positive link between pragmatic ability and cognitive factors such as EC or ToM. We discuss whether the lack of robust evidence in favour of a multilingual advantage in comprehending implicatures could be due to the bilectal status of our control group.

1. Introduction.

As a result of globalization, societies around the world are becoming increasingly more diverse and multilingual. Some surveys suggest that more than half of the world's population speaks more than one language in everyday life (Grosjean 2010) and that up to two thirds of children in the world are currently raised in bilingual environments (Crystal 1997). Bilinguals, therefore, make up a significant portion of the world's population while childhood bilingualism seems to be the rule rather than the exception in many places of the world. In this context the question of bilingual children's cognitive and linguistic skills relative to their monolingual peers has attracted the attention of a large number of researchers, especially over the past 25 years (Adesope et al. 2010). However, little is known about the effect of bilingualism on children's conversational understanding, an ability which is generally considered to be founded on a different type of knowledge -principles of social interaction- than linguistic competence. In this paper we aimed to investigate whether speaking more than one language confers an advantage for children's conversational abilities.

In the next sections we summarise the findings of previous researh on the effect of bilingualism on children's language and cognitive development (section 1.1.), we sketch the theoretical framework on which our investigation of bilingual children's conversational abilities is based–Grice's theory of conversation- (1.2.) and we briefly discuss studies that previously addressed the same question and consider their limitations (1.3.). We also examine theoretical and empirical reasons that suggest a functional link between conversational ability and executive control (1.4.). Finally, we briefly describe the language situation in the Republic of Cyprus where the current study was conducted (1.5.). Having motivated our research in this way, we then present our novel study on the effect of bilingualism on children's conversational development.

1.1. The effect of bilingualism on children's linguistic and cognitive development.

Research on bilingual children's linguistic and cognitive development has by now revealed two major trends. First, a large body of evidence shows that bilingualism has a negative effect on some aspects of children's linguistic performance. Bilingual children are typically found to possess smaller vocabularies in each of their languages when compared to their monolingual counterparts (as indicated by both productive and receptive measures of vocabulary size), although this disadvantage might disappear when considering conceptual vocabulary or total vocabulary (Junker & Stockman, 2002; Oller & Elliers, 2002 Pearson, Fernández, & Oller, 1995; Petitto & Kovelman, 2003; Poulin-Dubois et al. 2012). For instance, using a large sample of 1738 bilingual and monolingual children between 3–10 years old, Bialystok et al. (2010) demonstrated that bilingual children had a lower receptive vocabulary score than monolingual children at all age groups tested, although this bilingual vocabulary gap was evident only for words used in a home context but not for words used in the context of school. Similar results have been reported for bilingual children's grammatical knowledge. Studies comparing bilingual and monolingual children have shown similar delays in the development of certain morphological and syntactic structures for bilingual children (Gathercole 2002a, 2002b, 2002c; Chondrogianni & Marinis 2010).

Contrary to this pattern, recent research has reported beneficial effects of bilingualism for several aspects of children's non-linguistic cognitive functioning. There is now considerable evidence that the regular use of two languages by bilingual children enhances their executive control abilities (e.g. Bialystok 1999; Carlson & Meltzoff 2008; Martin-Rhee & Bialystok 2008; Poarch & van Hell 2012; Bialystok 2011; Morales et al. 2013; Bialystok 2010; Bialystok et al. 2009; though see Hilchey & Klein 2011; Morton & Harper 2007). Executive control (EC) refers to a heterogeneous cluster of higher order cognitive processes that underlie flexible, goal-directed behaviour (Mazuka et al. 2009). A widely accepted view proposed by Miyake et al. (2000) suggests that EC comprises three major processes: cognitive flexibility (the ability to flexibly switch attention between rules, representations, or perspectives), working memory (the ability to simultaneously maintain and manipulate task-relevant information in mind) and inhibition (the ability to suppress dominant, automatic, prepotent responses and resolve conflict by suppressing irrelevant information).

Studies investigating the effect of bilingualism on these skills in children have mostly reported a bilingual advantage in tasks requiring conflict resolution and inhibition (Bialystok et al. 2005; Carlson & Meltzoff 2008; Martin-Rhee & Bialystok 2008; Poarch & van Hell 2012; Bialystok & Viswanathan 2009; Bialystok 2010; de Abreu et al. 2012; Yang et al. 2011). It has been also suggested that the bilingual advantage in inhibition appears only under conditions that require conflict resolution but not in situations that require refraining from the execution of a salient or habitual motor response (Martin-Rhee & Bialystok 2008; Carlson & Meltzoff 2008). Recent studies have also provided some evidence for advanced cognitive flexibility (e.g. Foy & Mann 2013; Bialystok 2010; Bialystok & Martin 2004; Bialystok 1999; Barac & Bialystok 2012) and working memory abilities (Morales et al. 2013) in bilingual children. However, the effect of bilingualism on these skills is yet unclear as only a few studies have been so far reported and with sometimes mixed results (de Abreu 2011; Bonifacci et al 2011).

According to Bialystok (e.g. 2001) the bilingual advantage in inhibitory control is a consequence of the coexistence of two different languages in the bilingual mind/brain: because both language systems remain active when using one of them, bilinguals have to selectively attend to the relevant language and inhibit attention to the other language. This constant experience in selecting between two active conflicting language systems enhances bilinguals' inhibitory control mechanisms making them more efficient for other functions that extend beyond the linguistic domain. It should be noted, however, that the finding of a bilingual advantage in inhibitory control has not always been replicated while several researchers have questioned both the validity of the results of previous studies and the explanatory mechanism offered to explain this effect.

Typically, interference tasks (such as the Simon task or the Attentional Networks Task) require participants to provide responses to two types of experimental trials: incongruent trials where conflicting task-irrelevant information is presented and congruent trials where no interfering information is present. The interference effect calculated as the difference in reaction times between responses in incongruent and congruent trials is the standard measure taken from these tasks and is considered to be the main indicator of participants' inhibitory control abilities. However, in an empirical review of studies investigating the effect of bilingualism on performance in non-linguistic interference tasks, Hilchley & Klein (2011) concluded that empirical findings of a bilingual advantage on interference effects have been conspicuously absent for children and young adults. A more robust finding of this literature is that bilinguals show a global advantage in reaction times in both congruent and incongruent trials of interference tasks. They suggest that the finding of an overall advantage in RTs is simply not consistent with the hypothesis of a bilingual advantage in inhibitory control. If inhibitory control was the cognitive process affected by bilingualism then one would expect faster performance only in incongruent trials of interference tasks and a smaller interference effect for bilinguals. Following other researchers (Costa et al. 2009) they propose that the bilingual advantage in interference tasks might be better characterised in terms of an enhanced ability to monitor conflict rather than to resolve conflict *per se.* This conflict-monitoring system is responsible for the detection of conflict which subsequently leads to the adjustment of the level of inhibitory control to aid its resolution. Finally, some researchers have raised the concern that the bilingual advantage on interference effects or on global RTs might be due to other factors (e.g. socioeconomic status), not related to bilingualism, that have not been adequately controlled in previous studies. Morton & Harper (2007), for instance, argued that differences in bilingual and monolingual children's inhibitory control might have been due to uncontrolled demographic factors such as ethnicity and socioeconomic status. When directly controlling for these factors Morton & Harper (2007) found no differences in bilingual and monolingual children's performance in the Simon task (but see Bialystok 2009). However, subsequent studies that also controlled for these factors still reported a bilingual advantage (e.g. Carlson & Meltzoff 2008; Poarch & van Hell 2012; Yang et al. 2011).

1.2. Grice's theory of conversation.

Philosopher Paul Grice (e.g. 1975) argued for a division between the content of an utterance that depends on the meanings of words and the structural relationships between them (semantics) and the aspects of an utterance that depend on context-sensitive inferential processes (pragmatics). He suggested that a good deal of what is communicated relies on the appreciation of certain conversational expectations. In this regard speakers design their utterances with respect to the cooperative principle and maxims. Hearers, on the other hand, expect speakers to adhere to these conversational principles and calculate speaker's meaning on the basis of this expectation and the utterance's conventional meaning.

Grice's conversational maxims (Grice 1975; 1989) exhort interlocutors to *be no less and no more informative than is required for the purpose of the talk exchange* (quantity maxims I and II), *tell the truth and avoid statements for which they have not adequate evidence* (maxim of quality), *be relevant* (maxim of relation) and *be brief, orderly and avoid ambiguity and obscurity* (maxim of manner). Grice (1975) also proposed that other maxims might need to be developed in order to provide a full account of pragmatic meaning such as a maxim of politeness *be polite*. These conversational maxims form the basis of pragmatic competence which allows effective communication in context. Grice never intended his ideas to be connected to theories of cognitive processing. However, pragmatic theorists who have followed up on Grice such as Relevance theorists (e.g. Sperber et al. 2002) and neo-Griceans (e.g. Levinson 2000) share a cognitive-oriented perspective and aimed to provide an explanatory account of the mechanisms that deliver these different aspects of meaning and of how these mechanisms develop in use.

In many communicative situations consideration of the cooperative principle and maxims prompts interlocutors to draw inferences by which they attribute to speakers an implicit meaning that goes beyond what they literally said. These inferences are what Grice called *conversational implicatures*. An example is given in the following dialogue:

(1) John: Did all of the students pass the exam?
Mary: Some of the students passed the exam.

According to standard semantic accounts the literal meaning of Mary's utterance in (1) is compatible with the interpretation in (2), however through the conversational maxim of quantity what Mary actually conveys is the meaning in (3):

(2) Some and possibly all of the students passed the exams.
(3) Not all of the students passed the exams.

The proposition in (3) is an inference known as a scalar implicature. According to Grice's pragmatic theory drawing the implicature in (4) is a process that involves at least two different inferential steps: step I requires sensitivity to the maxim of quantity I (sensitivity to the fact that there is a more informative proposition using the term *all* that could have been used but wasn't) and step II involves the negation of the more informative proposition that leads to the implicature. In a similar vein, a broad class of inferences can be communicated and inferred by exploiting the other maxims.

1.3. Bilingualism and conversational understanding.

Three studies conducted by Michael Siegal and colleagues directly explored the extent to which bilingualism confers an advantage on preschool children's conversational understanding. Siegal et al. (2007) sought to determine whether 4- to 6-year-old bilingual children outperform their monolingual peers in the ability to understand scalar implicatures. Children were introduced to a puppet who often (but not always) said silly things. In the critical test trials the puppet used under-informative utterances (e.g. *Teddy put some of the hoops on the pole*) to describe situations in which a more informative utterance was appropriate (e.g. Teddy put all of the hoops on the pole). Children were asked to judge whether the puppet's utterances were correct. Rejection of the critical under-informative utterances indicated the generation of scalar implicatures. The results of their study showed that bilingual children were more advanced than their monolingual peers in computing scalar implicatures.

In two subsequent studies Siegal et al. (2009, 2010) employed a Conversational Violations Test (CVT) in order to examine whether bilingual children also excel in their ability to detect violations of Gricean maxims (quantity, quality, relation, manner and politeness). In the CVT children were presented with two doll speakers who answered a question. One answer violated a conversational maxim while the other was pragmatically appropriate. For example, for the maxim of quality, the question was *Have you seen my dog?*, the pragmatically inappropriate answer was *Yes, he's in the sky* and the pragmatically felicitous reply was *Yes, he's in the garden*. Children were asked to indicate the doll that said something silly. Siegal et al. (2009) found that bilingual children performed significantly better than their monolingual counterparts in the CVT. Similar results were obtained in their (2010) study in which they directly controlled for socioeconomic status (experiment 1), compared bilingual children's CVT performance in each of their languages to that of monolingual speakers of that language and included a group of children that were bilingual in a different pair of languages than previously tested (experiment 2).

Overall, the three pioneering studies by Siegal et al. (2007; 2009; 2010) provided compelling evidence for precocious development of conversational abilities in bilingual children. Similar advantages for bilingual children have been also reported for other aspects of their socio-cognitive development. Goetz (2003) and Kovacs (2008) reported evidence that bilingualism boosts children's Theory of Mind-the ability to understand other people's mental states and intentions. They showed that bilingual children outperformed their monolingual peers in false-belief tasks, a standard measure of Theory of Mind for preschool-aged children. Kovacs (2008) also provided evidence that this early consolidation of TOM in bilinguals is connected to their advanced executive control abilities. In a similar vein Genessee et al. (1975) showed that bilingual children were more sensitive to their listeners' communication needs than monolingual children when explaining how to play a game to a blindfolded and a non-blindfolded interlocutor. Their results indicated that bilingual children gave more information about the game's materials to the blindfolded listener than the monolingual children. Finally, results from three studies conducted by indicated that bilingualism fosters young children's ability to understand and use nonverbal referential gestures in order to infer a person's referential intent.

We believe, however, that the results of the three studies by Siegal and colleagues (2007; 2009; 2010) which directly tested for a relation between bilingualism and conversational understanding, were limited in two important ways. First, Katsos et al. (2011) argued that the CVT employed in Siegal et al. (2009; 2010) falls short of being a real test of children's pragmatic ability, at least as far as the maxims of quality and relation are concerned. As Katsos et al. (2011) note, given the question *Have you seen my dog?*, children can reject an answer such as *Yes, he's in the sky* just by drawing on their encyclopaedic knowledge about the world and not because they are sensitive to the maxim of quality. Finally, the CVT is a test of fairly simple pragmatic abilities -children are just required to detect the utterance that violates a conversational maxim. On the contrary, according to pragmatic theory, inferring implicatures requires a more sophisticated and complex pragmatic understanding which includes, but is not exhausted to, the detection of a violation of a conversational maxim. Thus even if the CVT were to measure some form of pragmatic ability, that of sensitivity to violations of Gricean maxims, it is still an empirical question whether bilingualism leads to an advantage in children's comprehension of implicatures –the kind of pragmatic inferences that interlocutors rely on in communication.

1.4. The relations between bilingualism, EC and conversational understanding.

Several theoretical and empirical factors suggest that EC might underpin the development of children's ability to comprehend implicatures. Firstly, according to pragmatic theory, successful generation of implicatures requires a complex reasoning process that takes into consideration a rich array of linguistic and extralinguistic information (literal meaning, the interlocutor's perspective, norms of conversation, background information). This requirement to coordinate different pieces of information predicts that the process of generating implicatures might depend upon executive resources and that this process is more difficult for populations (e.g. children, autistic individuals) whose resources are more limited than those of neuro-typical adults. Indeed Pouscoulous et al. (2007) showed that simplifying the experimental task and its linguistic material resulted in higher rates of scalar implicatures among children. As they argued, children saved precious cognitive resources that they in turn used to generate the implicature. Secondly, data from special populations like autistic people and bilingual children suggest a possible positive relation between the two skills: in the former case EC deficits co-occur with difficulties in conversational understanding while in the latter case advanced EC abilities co-occur with advanced conversational ability (e.g. Hill 2004; Ozonoff 1997; Loukusa & Moilanen 2009; Siegal et al. 2007; 2009; 2010).

Thirdly, findings from the developmental literature on children's Theory of Mind abilities suggest that aspects of EC such as working memory and inhibition are implicated in the expression and emergence of these skills in children (e.g. Moses et al. 2005). This is particularly relevant to pragmatic competence as comprehending implicatures involves a mind-reading process about the speakers' intentions and thus depends on ToM abilities (e.g. Sperber & Wilson 2002). In this respect if a functional link exists between ToM and EC, then it is possible that this link extends to pragmatic competence as well. Indeed recent research on children's (Nilsen & Graham 2009) and adult's (Brown-Schmidt 2009) perspective-taking abilities during referential communication provided evidence for a relation between the ability to take another's perspective and inhibitory control mechanisms. Nilsen & Graham (2009), for instance, found that children who possessed less developed inhibitory control skills were more prone to consider or even choose an object that was in their privileged ground and matched the speakers' referential description rather than to take into account their interlocutor's perspective. Finally, there is experimental evidence to suggest that the processing of scalar implicatures in adults involves working memory resources. De Neys & Schaeken (2007) and Dieussaert et al. (2011) reported that burdening adults' working memory resources with the requirement to remember a complex dot pattern before judging under-informative utterances decreased the rate of scalar implicature responses. In a similar vein, Antoniou & Katsos (2011) found that the rate of scalar implicature responses to under-informative utterances was significantly predicted by working memory capacity.

The three studies by Siegal and colleagues (2007; 2009; 2011) also explored the extent to which bilingual children's advantage in pragmatics is due to their enhanced EC mechanisms. In their 2007 and 2009 studies all children were tested on two measures of EC, the Day/Night task and the Card Sort Task, while in their 2010 study only the Day/Night task was administered. No significant relations between children's pragmatic abilities and EC were established. However, we argue, that there are several problems with both the pragmatic and the EC tasks used by Siegal and colleagues that limit the conclusions we can draw from their studies with regards to this issue.

Firstly, the studies by Siegal et al. (2007; 2009; 2010) were limited in that they employed only two measures of EC (the Card Sort task and the Day/Night task) and did not include any measures of working memory. Secondly, the selection of the EC tasks in their studies was not eventually felicitous in several respects: for the Card Sort task, Siegal and colleagues (2007; 2009) reported a near ceiling level of performance in both studies in which this test was used. With regards to the Day/Night task it has been suggested that the bilingual advantage in inhibition is mostly found in tasks which require conflict resolution (such as the Simon task) but not in measures of response inhibition (such as the Day/Night task). No measures of the former cognitive ability were included in their studies. Indeed the selection of these specific tasks might have actually masked potential differences between the two groups and, consequently, potential relations between conversational ability and EC. Finally, to reiterate a point made above, the CVT employed in their studies (2009; 2010), is a test of fairly simple pragmatic abilities which one might not actually expect to correlate with measures of EC. On the contrary, inferring implicatures is typically found to be a costly process (e.g. Bott & Noveck 2004; DeNeys & Schaeken 2007) and one could predict that this ability is more dependent on executive resources.

1.5. The language situation in Cyprus.

The present study took place in (the Republic of) Cyprus where the population is mostly Greek Cypriot. The country's linguistic profile is typically described as diglossic with Cypriot-Greek (CG) as the low variety and Standard Modern Greek (SMG) as the high variety (Arvaniti 2010; Grohmann & Leivada 2012; Rowe & Grohmann under review). CG is an oral, not orthographically codified, variety traditionally characterized as a dialect of Greek (Newton 1972; Kontosopoulos 2001; Arvaniti 2010). It is natively acquired and used for everyday communication. SMG, on the other hand, is the official, constitutionally recognized[2], language of the Republic of Cyprus. It is learnt mainly through formal education (from kindergarten onwards), it is used in all forms of formal writing and it is the language of the media. CG and SMG show substantial overlap in both grammar and lexicon although they are also known to differ in several respects: morpho-syntax (e.g. clitic placement, formation of *wh*-questions), phonetics (e.g. prenasalization of plosives in CG), phonology (replacement of SMG [j] with [i] as in [sa.va.to.ci.ri.a.ko] instead of [sa.va.to.ci.rja.ko]) and in the use of certain lexical terms (Arvaniti 2010). Following Rowe & Grohmann (under review) we use the term *bilectalism* to describe the linguistic status of

[2] The other official and constitutionally recognized, language of the Republic of Cyprus is Turkish.

people speaking CG and SMG in Cyprus.

Until very recently, SMG was by law the language of instruction at schools and CG was not taught at any level of education. However, with the new "Curriculum for the Modern Greek Language", whose implementation started in September of 2011 teachers are encouraged to use the CG dialect in the classrooms and CG becomes an object of instruction (Hadjioannou & Tsiplakou 2011). The parallel use and contrastive analysis of CG and SMG in the classroom is expected to enhance the pupils' metalinguistic knowledge at both the grammatical and communicative-sociolinguistic level (Hadjioannou & Tsiplakou 2011). The new curriculum is of particular relevance to our study as the present study was conducted after the start of its implementation.

1.6. The present study.

The aim of this study was twofold: The first aim was to establish whether multilingualism confers an advantage in children's ability to comprehend implicatures. The second aim was to investigate whether a potential multilingual advantage in pragmatics is mediated by some aspect of their EC or ToM abilities. In order to achieve these aims we tested multilingual and bilectal children growing up in the Republic of Cyprus. Children were administered (a) an extensive battery of tasks that cover all aspects of EC (cognitive flexibility, working memory and inhibition) and ToM and (b) a novel and extensive conversational test examining their ability to understand a wide range of implicatures based on Grice's maxims of conversation.

2. Method.

2.1. Participants

There were 66 children in the study, consisting of 36 bilectal children (in CG and SMG; 18 boys and 18 girls; ages 6.6–12.3, mean age 8.5, SD 1.39 years; see table 1) and 30 multilinguals (bilectals in CG and SMG, also speaking one or more additional languages; 14 girls and 16 boys; ages 6.6–11.8, mean age 8.8, SD 1.58 years). Background characteristics of the two groups are presented in table one. All children were recruited from schools in the Republic of Cyprus. Miltilingual children were recruited from two private schools offering English-speaking programs. As the language of instruction was not SMG, these schools did not follow the national curriculum, although they included SMG as a separate subject. 33 of the bilectal children were recruited from a private school while the rest were recruited from a local public school. Both of these schools offered traditional Greek-speaking programs with limited exposure to a second language. Parents were given a Language Background Questionnaire to examine children's language exposure. In the multilingual group 22 of the children were reportedly exposed to both CG and English from birth, four of the children were exposed to English from 12 months of age, one from 36 months and one from 48 months of age. For the six children who were not exposed to English from birth it was reported that one of their parents had a place of birth other than Cyprus and usually or always used a language other than CG with the child. For four of these children it was reported that they were exposed from birth to a third language other than CG and English. For 27 of the multilingual children it was reported that for each language at least one of their parents usually or always used that language when speaking to the child or that parents used approximately equal proportions of both languages when speaking to the child. Four of the multilingual children had parents who only sometimes or rarely used CG when directing speech to their child. However, for these children parents indicated that, generally, they were usually or always exposed to CG in the community. Nine additional children were not included in the analyses because either they were recruited as multilinguals and their language background questionnaires indicated that they had limited exposure to a second language (seven children) or they were recruited as bilectals and their questionnaires indicated that they were extensively exposed to a second language (two children).

2.2. Materials and procedure.

All children were given (1) a novel conversational test and (2) a battery of cognitive tests potentially relevant to implicature generation. Children were tested in two sessions taking approximately 50 minutes each.

2.2.1. Conversational test.

This was a test of children's pragmatic abilities testing comprehension of relevance, manner, and scalar implicatures, as well as metaphor and irony, which according to the Gricean account involve the maxim of quality. There were two versions of the conversational test and each child was tested in one of the two versions. For the two scalar implicatures sub-tests (described in more detail below) the same items were used in both versions. The actual stories or sentences used were in CG and were prerecorded by a CG native speaker. At the beginning of the test children were presented with a picture of a character whose name was George. They were informed that George was a boy of their age and that they would hear stories about George or sentences uttered by George himself. Each sub-test is described in more detail in the sections that follow.

2.2.1.1. *Relevance Implicatures.*

This was a test of children's ability to comprehend implicatures that are based on Grice's maxim of relevance (*Be relevant*). It was a story completion task based on a previous study with children conducted by Bernicot et al. (2007). Children were instructed that they would hear stories about George and his mother and that at the end of each story they had to point to a picture that showed how the story ended. There were three test items. Each item was composed by two slides. In the first slide children were shown a picture that visually established the story's interaction setting and the target story was heard. Target stories in the three test items were of the following format: George asked his mother a question (e.g. *Mom, can I buy an ice-cream?*) and his mother replied with an utterance that implied either a negative or a positive answer (*You are ill* or *I have money in my wallet* respectively). In the second slide the experimenter introduced two pictures as possible endings to the target story. The experimenter asked the child *What happened at the end of the story*? and proposed the two pictures as possible endings to the story (e.g. *George bought an ice-cream* and *George did not buy an ice-cream*). One of the pictures depicted how the situation would be if the reply to George's question had been positive (e.g. George eating an ice-cream) and the other showed how the situation would be if the reply had been negative (e.g. George doing something else such as playing with his toys). There were two versions for each story: one version of the story ended with an utterance that implied a negative response and the other with an utterance that implied a positive response. This also assured that both pictures in the second slide of the test item would be chosen by some children. A sample test item of this sub-test is given in the appendix (pictures one and two).

2.2.1.2. *Metaphor.*

This was a test of children's ability to understand metaphors which according to Grice's pragmatic theory (Grice 1989) are based on the maxim of quality (*Do not say what you believe to be false*). This test was designed based on two previous studies with children conducted by Waggoner and colleagues (1989; 1997).

Children were instructed that they would hear stories about George and his father and that at the end of each story they had to point to a picture that showed how George's father felt at the end of the story. They heard three stories ending in metaphors describing emotions of sadness (e.g. *When he returned back home George's father was a melting snowman*) or anger (e.g. *When he returned back home George's father was a thundering cannon*). Metaphors describing emotions were chosen because as Waggoner et al. (1989:132–133) note there is evidence that even preschool children have knowledge about emotions, that they use them in their spontaneous speech and can differentiate simple emotions such as anger, sadness, happiness, fear, love or hate (Bretherton et al. 1986; Stein & Levine 1987; Borke 1971; Smiley & Huttenlocker 1984; Schwartz & Trabasso 1984). Unfamiliarity with the semantic domain to which a metaphor refers has been argued to be one of the factors that contribute to children's poor performance with metaphors in previous studies (Keil 1986; Winner 1988; Vosniadou et al. 1984). Indeed Waggoner et al. (1989; 1997) showed that even 5-year-old children can successfully interpret psychological metaphors describing emotions at above chance levels.

We also chose to contrast two negative emotions, the emotions of sadness and anger, so that to be fairly sure that children who successfully select the correct picture will do so because they in fact interpret the metaphor. If, for example, one was to contrast a negative and a positive emotion such as the emotions of sadness and happiness, then children selecting the correct picture (the picture of the sad man) when presented with a metaphor such as *George's father is a melting snowman* cannot unambiguously be granted with a successful understanding of the metaphor. This is because one can argue that they in fact select the correct picture simply on the basis of matching a negative event as described by the metaphor vehicle (a snowman that is melting) with a negative emotion (sadness) instead of actually interpreting the metaphor (see Waggoner et al. 1989). Waggoner et al. (1989) used similar stories to the ones used in this test and found that 5-year-old children could successfully comprehend metaphors describing emotions even when they had to decide between two emotions both of which were positive or both of which were negative. Nevertheless, their results showed that 5-year-old children's and first graders' performance was not above chance when novel metaphors were used (as in our study).

As in the relevance implicatures sub-test each test item was composed by two slides. In the first slide children were presented with a picture relevant to the story and the target story was heard. In the second slide the experimenter asked the child *How did George's father feel at the end of the story?* and introduced two pictures: a picture of a sad man and a picture of an angry man. All metaphorical expressions were embedded in contexts that introduced the two emotions but were neutral with regards to the emotion expressed by the metaphors. This was achieved by including a sentence such as the following in all story contexts: *George was thinking of how his father would feel when coming back home. He didn't know whether his father would feel sad because his favorite team lost or whether he would feel angry because the team did not play well*. Finally, only novel and apt metaphors were used. 15 metaphors generated by the experimenters were given to 30 adult native speakers of Cypriot-Greek. For each metaphor it was indicated which emotion it expressed and participants were asked to rate it for aptness and novelty on a 5-point scale. Only metaphors with a mean rating above 2.5/5 for both novelty and aptness were used in this experiment. Metaphors were distributed and rated by participants using the internet software Qualtrics (http://qtrial.qualtrics.com). A sample test item of this sub-test is given in the appendix (pictures three and four).

2.2.1.3. *Irony.*

This was a test of children's ability to comprehend ironic utterances which according to

Grice's pragmatic theory (Grice 1989) are based on the maxim of quality. As with relevance implicatures this was a story completion test based on the study by Bernicot et al. (2007). Children were instructed that they would hear stories about George and that at the end of each story they had to point to a picture that showed how the story ended. There were three test items ending in ironic utterances and three filler items ending in literal utterances. In the three test items stories were of the following format: George asked a question (e.g. *Shall I turn on the heater?*) and a second character replied with an ironic utterance (*No, I'd prefer to catch a cold*) that implied the opposite of what was explicitly said. All ironic utterances implied a positive response to George's question. There were two cues indicating that the final target utterance must be interpreted ironically: firstly the story contained information that was discrepant with a literal interpretation of the final target utterance (e.g. *Mary was really cold and wanted something to warm her up*) and secondly the final utterance was uttered in an ironic, mocking intonation. In filler items stories were of a similar format with the difference that they ended in literal utterances that were consistent with the contextual information provided in the story. Filler items were included so that the interpretation of the final target utterance would not always be the opposite of what was said. In filler items the final target utterance was either a negative or a positive response to George's question.

Each item was again composed by two slides. In the first slide children were shown a picture that visually established the story's interaction setting and the target story was heard. In the second slide the experimenter introduced three pictures as possible endings to the target story. The experimenter asked the child *What happened at the end of the story?* and proposed the three pictures (e.g. *George turned on the heater*, *George did not turn on the heater* and *George brought Maria a coat to wear*). One picture was compatible with an ironic interpretation of the final target utterance (e.g. *George turned on the heater*), a second picture was compatible with a literal interpretation of the final target utterance (e.g. *George did not turn on the heater*) and a third picture was compatible with an interpretation that relied on the contextual information provided in the story but not with the target utterance itself (e.g. *George brought Maria a coat to wear*). There were two versions for each story that differed only with regards to George's contextual question: in the first version, for instance, George would ask the second character *Shall I turn on the heater?* while in the second version the question was changed to *Shall I bring you a coat to wear?*. This assured that all pictures in the second slide of the test items would be chosen by some children: for the example given above, for instance, pragmatic responders in the first version of the story would choose the first picture, pragmatic responders in version two of the story would choose the second picture while literal responders in both versions would choose the third picture. A sample test item of this sub-test is presented in pictures five and six of the appendix.

2.2.1.4. *Manner implicatures.*

This was a test of children's ability to comprehend implicatures which according to Grice's pragmatic theory are based on the maxim of manner (*Be Clear*) and more specifically the sub-maxim *Be brief (avoid unnecessary prolixity)*. The general format of this test was a sentence to picture matching task. Participants were instructed that they would hear George describing a picture from a book and that they had to point to a picture that matched George's description. There were six items in total, of which three were critical and three filler items. Critical items were causatives for which a lexicalized and an opposed periphrastic alternative are available (e.g. *In this picture a man opened the door* as opposed to *In this picture a man made the door open*). Lexicalized causatives are associated with a direct, normal, more frequent, stereotypical causation while their periphrastic alternatives are associated with an indirect, non-normal, less frequent, non-stereotypical causation (Levinson 2000). Filler items were literal expressions for which the correct picture could be identified on the basis of their explicit content. Each item was composed by two slides: the first slide displayed an image of George holding a book and the target sentence was heard. The second slide displayed two pictures as possible matches to George's description. In the critical items the two pictures contrasted an unmarked, normal, stereotypical way of causation (e.g. a picture of a man opening the door using a key) with a marked, non-normal, non-stereotypical way of causation (e.g. a picture of a man opening the door by kicking it). There were two versions for each item: in the first version the unmarked, lexicalized expression was used while in the second version its corresponding marked, periphrastic expression was used. A sample test item of this sub-test is given in the appendix (pictures seven and eight).

2.2.1.5. *Scalar implicatures.*
This was a test of children's ability to understand implicatures based on Grice's first maxim of quantity (*Make your contribution as informative as is required (for the current purposes of the exchange)*). This test was a PowerPoint version of the action-based task used by Pouscoulous et al. (2007). Participants were presented with slides depicting five boxes and a selection of animals (five elephants, five turtles, five dolphins and five hippopotamus). As in the Pouscoulous et al. (2007) study, there were three scenarios: in the 5/5 scenario all boxes contained the same kind of animal (e.g. all boxes contained one elephant each), in the 2/5 scenario two of the five boxes contained the same kind of animal (e.g. two boxes contained a turtle each, the rest having no animals). For each scenario children heard statements constructed with the quantifiers *all*, *some* and *none* and three kinds of animals (elephants, turtles and dolphins). This resulted in a total of 27 test items in the study. There were also three practice items in order to familiarize participants with the format of this task. Children were instructed that they would hear George describing the display and they had to make the display match George's description using the mouse. There were three critical items. These were statements with the quantifier *some* (e.g. *There are turtles in some of the boxes*) in the 5/5 scenario (e.g. in which all five boxes contained a turtle). If children were sensitive to the maxim of quantity and calculated the scalar implicature that *some* implies *some but not all* then they should have removed at least one (but not all) of the animals from the boxes. A sample critical item is given in picture nine of the appendix.

2.2.1.6. *Scalar implicatures binary judgment task.*
This was again a test of children's ability to comprehend scalar implicatures. It was designed and administered using e-prime psychology software. In each trial of this test, the participant was shown a depiction of three or five cards, face down. A fixation cross was shown (500ms) and then an auditory stimulus was played, "There are <X> on <Q> of the cards", where X was the item type (rings, hearts, stars) and Q the quantifier (*all, some, not all* or *none*). When the auditory stimulus ended, the cards were immediately 'turned over' to reveal the items. Participants were instructed to press a green-labeled key if the utterance was true, and a red-labeled key if it was false, responding as quickly and accurately as possible. 48 trials were administered to each participant, comprising three blocks of 16 trials. There were three critical under-informative items each presented in one of the three blocks. The rest of the items comprised an equal number of true and fully informative utterances and semantically false utterances. A sample critical item is given in picture 10 of the appendix.

2.2.2. Working memory tests.

2.2.2.1. *The Backward Digit Span Task (Wechsler, 1949):*
In each trial of this task participants heard an auditory stimulus spoken by a native GG speaker presenting a list of numbers (e.g. *5, 6, 3*). Participants were instructed to repeat the numbers in reverse order (e.g. *3, 6, 5*). The task began with two digits and the string of digits

increased after two successful trials on the same number of digits. The task was discontinued when participants erred on two consecutive trials on the same number of digits (highest level was eight digits). There were two practice trials with a string of three digits before the actual test. One point was awarded for each successful trial. Scoring was conducted by an experimenter present in the testing room during the task.

2.2.2.2. *The Corsi blocks Task:*

This was an online version of the Corsi Blocks task (Corsi 1973). It was administered as a measure of visuo-spatial short-term and working memory. There were two parts of this test: the forward and backward condition. The two parts were administered separately with the forward condition always administered first. In each trial of this task participants were presented with a display of nine boxes. In the forward condition the boxes lit up in a specific order and participants were instructed to click on the boxes in the same order. In the backward condition the boxes lit up in a specific order and participants had to click on the boxes in the reverse order. Both conditions started with a sequence of two boxes lighting up and increased by one box after every second trial (highest level for forward condition was nine boxes while for backward condition was eight boxes). There were two practice trials with a sequence of two boxes for each condition. One point was awarded for each successful trial.

2.2.3. Inhibitory control tests.

2.2.3.1. *The Soccer Task:*

This was an online version of the Stop-Signal task (adapted from Logan 1994). It was administered as a measure of response inhibition. In each trial of this task children were presented with displays showing two football pitches (one at the right side of the display and one at the left side) and a ball on one of the two pitches. They were instructed that their job was to indicate which direction to run for the ball. When the ball was on the right side they had to press the right arrow key on the keyboard and when the ball was on the left side they had to press the left arrow key on the keyboard. They were told that they had to respond as quickly and accurately as possible. On approximately 20% of the trials the children heard a referee blowing his whistle. They were instructed that when they heard the whistle blowing they had to stop and not press any buttons until the picture disappeared and moved on to the next display. There were three sets of 36 trials that included equal numbers of stimuli with the ball on the left-hand and right-hand sides. During each set about 20% of the trials were stop trials (i.e., a whistle blowing). Across the three sets, there were a total of 108 trials, with about 20% of those being stop trials and 80% being go trials. There were also 10 practice trials. At least two of these 10 trials were stop trials (i.e., a whistle blowing).

In stop-signal trials the time interval between the presentation of a stimulus and the emission of the stop signal cue (the whistle blowing) varied depending on the participant’s performance. This procedure ensured that participants will correctly inhibit a response in stop-signal trials approximately 50% of the time. When the time interval increased, it was more difficult to correctly inhibit a response. In this respect, if a participant was performing well (fast enough in the go trials and accurately in the stop trials), the time interval increased until a mistake was made. If the participant’s performance was poor the delay period decreased. The main dependent variable taken from this task was the Stop Signal Reaction Time (SSRT). This was calculated by subtracting the average delay period in stop-signal trials from the average reaction time of responses in the go trials. The task started with a delay at 250 ms which tended to decrease or increase by 50 ms depending on the participants’ performance.

2.2.3.2. *The Simon Task (Simon, 1969):*
This task was administered as a measure of interference suppression. In this task participants were instructed to press the right arrow key on the keyboard if a red square appeared on the screen and the left arrow key if a green square appeared. In congruent trials each square appeared on the same side as the appropriate button to be pressed (e.g. a red square on the right of the screen). In incongruent trials the square appeared on the opposite side to the appropriate arrow key (e.g. a red square on the left of the screen). This created a conflict between an irrelevant feature of the stimulus that had to be inhibited (the square's position on the screen) and a relevant feature that determined the correct response (the square's color). In neutral trials the red or green square appeared at the center of the screen. The task included two blocks of trials presented in a fixed order. In the first block 48 congruent and 48 incongruent trials were randomly intermixed whereas the second block included only 48 neutral trials. In both blocks the actual test trials were preceded by eight practice trials. The main dependent measure was the difference in mean reaction times between congruent and incongruent trials in the first block (Simon effect).

2.2.4. Cognitive flexibility test.
The Color-Shape Task:
This was an online version of the Figure Matching Task used by Ellefson et al. (2006). In each trial of this task participants were presented with a display that comprised several simultaneous stimuli. The first was the target figure that was presented in the center of the display. The target figure could be one of two shapes (triangle or circle) and one of two colors (blue or red). Secondly, two small figures were shown at the bottom of the display, one on the left and one on the right. One of these small figures was a triangle and the other was a circle. The small figures could be one of two colors: one was red and the other was blue. Finally, the cue to the matching task was presented at the top of the display, just above the target figure. The cue was either the word *color* flanked by two small green Xs or the word *shape* flanked by two small squares. Both words were written in English. Due to the fact that we aimed to compare performance of bilectal children who had little exposure to English and multilingual children who had extensive exposure to English we decided to occlude the words (*color* or *shape*) from the matching cue in the display using a sticker. This was done to eliminate potential differences in performance between the two groups that might have resulted due to differences in their knowledge of these words in English (although we could be fairly sure that the multilingual children knew the words *color* and *shape* in English we couldn't be so sure about the bilectal children). Children were instructed to select one of the small figures at the bottom of the screen that matched the target figure based on the matching cue at the top of the screen. If the matching cue was the two small squares, they had to select the small figure that matched the target figure for shape (shape game) and if the matching cue was the two small green Xs, they had to select the small figure that matched the target figure for color (color game). In order to select the small figure at the bottom left side of the display they had to press the left arrow key and in order to select the bottom right small figure they had to press the right arrow key on the keyboard. Children were told that they had to respond as quickly and accurately as possible. The experiment comprised four blocks which appeared in random order: a pure color and a pure shape block and two mixed blocks. The pure color and pure shape blocks did not require children to switch between the two games and children had to perform only the color game or only the shape game respectively. The two mixed blocks included switches between the two games every two trials. There were two types of trials: repeat trials and switch trials. Repeat trials occurred when children repeated the same game as the previous trial while switch trials occurred when participants changed to a different game from the previous trial. Repeat trials occurred in both pure and

mixed blocks whereas switch trials occurred only in the mixed blocks. We took three main dependent variables from this task. The first was the mixing cost calculated as the difference in mean reaction time between correct repeat trials in the pure blocks and all correct repeat trials in the mixed blocks. The second dependent measure was the switching cost. This was calculated by subtracting mean reaction time in correct repeat trials from mean reaction time in correct switch trials in the mixed blocks. Finally we also calculated an efficiency score that took into consideration both accuracy and mean reaction time. This score was calculated by dividing the percentage of accurate responses with mean reaction time (in seconds). A separate efficiency score was calculated for pure and mixed blocks.

2.2.5. Theory of Mind test.

Frith-Happe Animations Task (White et al. 2011):

In this task children were presented with videos of geometric animations (a big red triangle and a small blue triangle) moving around the screen. There were 12 videos divided into three types of animations: random, goal-directed and theory of mind. In the random set of animations the two triangles were randomly moving around the screen without any interaction between them and without any purpose. For instance, the two triangles were presented bouncing off the walls like billiards balls or drifting about. In the goal-directed set the two triangles were doing something together (interacting) in order to achieve specific goals. The action of one triangle affected the physical action or behavior of the other. For instance, the two triangles were dancing or fighting. Finally the ToM animations, presented the two triangles interacting so that one triangle manipulated the feelings and thoughts of the other. For example, one triangle mocked or seduced the other. Children were asked to watch each video carefully and at the end categorize it as one of three types: *no interaction* (random), *physical interaction* (GD) and *mental interaction* (ToM). Following White et al. (2011) children were provided with the following definitions of each category: ***no interaction*****:** *there is no interaction between the two triangles and movement appears random. The movement of one triangle does not affect the movement of the other triangle. The movements of the two triangles appear disconnected from each other*; ***physical interaction:*** *an interaction between the two triangles in which the actions of one triangle affect the actions of the other triangle in order to achieve specific goals*; ***mental interaction*****:** *an interaction between the two triangles in which the actions of one triangle affect the feelings and thoughts of the other*. Children were also instructed that at the end of some videos the experimenter would ask some more questions about the animated triangles they had just watched. There were three practice videos (one for each type of animation) during which children were given detailed feedback and explanations in order to ensure they understood the task. At the end of each video in the experimental phase the experimenter presented the three categories, repeated their definitions and asked the child to categorize it. For each ToM video, after the categorization task, children were further asked two questions about the triangles' mental states. They were given a list of five adjectives and were asked to choose the one that best matched each triangle's feeling at the end of the video. The dependent measure in this task was the total number of correct categorizations for the ToM videos plus the total number of correct responses in the multiple choice questions about the triangles' feelings (ToM-MCQ-feelings).

2.2.6. Expressive vocabulary test.

Word Finding Vocabulary Test:

The standardised Greek version (Vogindroukas et al. 2009) of the Renfrew Word Finding Vocabulary Test (Renfrew 1995) was administered to assess expressive vocabulary. The test consisted of 50 black and white cards depicting objects. Children were asked to name the noun depicted on each card. Testing was discontinued after five consecutive wrong responses. Words from both CG and SMG were accepted as correct.

2.2.7. Test of non-verbal IQ.

The WASI Matrix Reasoning Test (Wechsler 1949):

This test was administered as a measure of general intellectual ability. In this test participants were presented with pictures depicting a matrix from which a section was missing. They were required to complete it by pointing at or stating the number of the correct response from five possible choices. Testing was discontinued after five consecutive wrong responses or after four wrong responses in five consecutive items.

2.2.8. Socioeconomic status measure.

The Family Affluence Scale (Currie et al. 1997):

This was a four-item questionnaire administered as a measure of the family's socioeconomic status. The questionnaire included the following items: *Does your family own a car, van or track? (No [0]; Yes, one [1], Yes, two or more [2]), Does your child have his own bedroom? (No [0]; Yes [1]), During the past 12 months, how many times did you travel away on holiday with your family? (Not at all [0]; Once [1]; Twice [2]; More than twice [3])* and *How many computers does your family own? (None [0]; One [1]; Two [2]; More than two [3]).* Based on the parents' answers, each child was given a score from 0–9.

3. Results.

3.1. Background measures.

Table 1 contains information about the background characteristics of the two groups. The two groups did not statistically differ for IQ, age and gender (r (two-tailed) =.136, p>.05, r(two-tailed)=.188, p>.05 and r(two-tailed)=.048, p>.05 respectively). However, multilingual children possessed a significantly smaller vocabulary than bilectal children (r(two-tailed)=-.760, p<0.001) and multilingual children had a higher FAS score than bilectal children indicating higher socioeconomic status (r (two-tailed)=.253, p<0.05). Nevertheless, neither FAS score nor vocabulary significantly correlated with any of the other dependent measures in the various tests used (all two-tailed *p*s >.05).

Table 1. Background characteristics of the monolingual and bilingual groups.

Group	N	Age	FAS	IQ	Vocabulary
Bilectals	36	8.3	5.94	15.74	34.51
		(1.78)	(1.78)	(6.30)	(4.88)
Multilinguals	31	8.9	6.90	17.61	17.94
		(1.55)	(1.70)	(7.49)	(9.20)

3.2. Conversational Test.

Means and standard deviations of bilectal and multilingual children's scores in each implicature sub-test out of a maximum correct score of 3 are shown in table 2. There was a ceiling effect of performance for relevance implicatures with the strict majority of children (n=63) responding correctly in all 3 items. Adequate variability was achieved in all other sub-tests. A 2 (language group: monolingual vs. bilingual) x 2 (version: 1 vs. 2) x 6 (type of implicatures: relevance, metaphor, manner, irony, scalars 1, scalars 2) mixed ANOVA with language group and version as between subjects factors and type of implicatures as a within subjects factor revealed the following results: a significant effect of version (F(1, 62)=11.517, p<.01) indicating that version 2 of the conversational test was more difficult than version 1; a tendency for an effect of group (F (1, 62)=2.585, p=0.113) indicating a numerical trend towards a bilingual advantage in overall performance in the conversational test; a marginal group x version interaction (F (1, 63)=3.726, p=.058) indicating a numerical trend towards a bilingual advantage in overall performance in version 1 of the test.

Mauchly's test indicated that the assumption of sphericity had been violated, W=.473, p<.0001, therefore, for the repeated measures effects and their interactions, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity (ε=.90). When looking at the within subjects effects, we observed the following results: a significant effect of type of implicatures (F (4.50, 62)=33.726, p<.001) indicating that certain types of implicature were more difficult for children. Post-hoc pairwise comparisons with Bonferroni correction for multiple comparisons revealed the following significant differences (all *p*s<.005): performance in the relevance implicatures sub-test was significantly different from all other implicature sub-tests indicating that relevance implicatures were easier to understand than all other types of implicature; performance in the irony sub-test was significantly different from all other implicature sub-tests indicating that irony was the hardest to understand for children. No other comparisons were significant. The type x version interaction was significant (F(4.50 , 62)=7.50, p<.0001) indicating that children performed significantly better in version 1 than in version 2 in comprehending metaphors (F(1, 64)=5.438, p<.05) and irony (F(1, 64)=22.471, p<.0001). The type x version x group interaction was also significant (F(4.50, 62)=2.475, p<.05). In order to further explore this interaction we ran separate ANOVAs for each one of the implicature types with version and language group as between subjects factors. These ANOVAs revealed the following results: for irony a significant effect of version (as above), a marginally significant effect of language group (F(1, 62)=2.612, p=.111) indicating a numerical trend towards a bilingual advantage and a significant language group x version interaction (F(1,62)=9.150, p<.005) indicating significantly better performance of multilinguals only in version 1; for scalars 2 a marginally significant effect of language group (F(1, 62)=2.827, p=.098) indicating numerically better performance for multilinguals; for manner implicatures a marginally non-significant language group x version interaction (F(1, 62)=3.155, p=.082) indicating that multilinguals were numerically better in version 1 of the test whereas bilectals were numerically better in version 2 of the test. No other effects or interactions were significant (all *p*s>.05).

Table 2. Mean scores (out of 3) and standard deviations of the monolingual and bilingual children on the six implicature tests.

Group		Relevance	Metaphor	Manner	Irony	Scalars 1	Scalars 2
Bilectals	Mean	2.94	2.11	2.19	0.78	2.44	2
	SD	(0.23)	(1)	(0.79)	(1.27)	(1.13)	(1.10)
Multilinguals	Mean	2.91	2.27	2.10	1.10	2.57	2.40
	SD	(0.40)	(0.91)	(0.80)	(1.42)	(0.90)	(0.97)

3.3. Executive control tests.

3.3.1. Simon Task.

Mean reaction times for correct responses in the Simon task are plotted in table 3 by condition and language group. We conducted a 3 (congruency: congruent, incongruent, neutral) x 2 (language group: monolinguals, bilinguals) mixed ANOVA, with congruency as a within subjects factor and language group as a between subjects factor. Mauchly's test indicated that the assumption of sphericity had been violated, W=.673, p<.0001, therefore, for the repeated measures effects and their interactions, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε=.73). The ANOVA results showed a significant effect of congruency ($F(1.41, 86.12)=26.951$, $p<.0001$) and a significant effect of language group ($F(1, 61)=6.151$, $p<.05$). Post-hoc pairwise comparisons with Bonferroni correction for multiple comparisons revealed the following results: incongruent trials (mean=963.433, SE=31.060) were significantly slower than both congruent (mean=888.928, SE=28.869) and neutral trials (mean=822.659, SE=20.942); neutral trials were significantly faster than both incongruent and congruent trials. The critical congruency x language group interaction was marginally significant ($F(1.41, 86.12)=2.911$, $p=.077$), indicating that multilingual children were faster in all conditions of the Simon task. However, a 2 (congruency: incongruent versus congruent) x 2 (language group: multilinguals versus bilectals) mixed ANOVA and a 2 (congruency: incongruent versus neutral) x 2 (language group: multilinguals versus bilectals) mixed ANOVA showed that the critical language group x congruency interactions ($F(1, 61)=4.239$, $p<.05$) and $F(1, 61)=4.117$, $p<.05$) were significant indicating that multilinguals were significantly faster than bilectals in incongruent trials than they were in congruent or neutral trials. Another way to consider whether multilinguals demonstrated better inhibitory control skills is to directly compare the Simon effect (calculated as the RT difference between incongruent and congruent trials) in the two groups. Indeed this score was significantly different for the two groups ($F(1, 61)=4.239$, $p<.05$), indicating better performance for the multilingual children.

3.3.2. Soccer Task.

Mean RTs and mean percentage of accurate responses in the go trials, mean delay, and mean SSRT for each language group are reported in Table 3. SSRT was the main dependent measure for this task. An ANOVA revealed no significant differences between the two groups for this score ($F(1, 61)=.340$, $p>.05$).

3.3.3. Color-Shape Task.

Mean RTs and mean percentage of accurate responses per block (pure versus mixed) and per trial type (repeat versus switch in mixed blocks) are reported for each language group in table 3. For each language group we also report the means for the following measures: switching cost (calculated as the difference in mean RTs between switch and repeat trials in mixed blocks), mixing cost (calculated as the difference in mean RTs between repeat trials in mixed blocks and repeat trials in pure blocks), efficiency score in the mixed and pure blocks (calculated by dividing the percentage of accurate responses by mean RTs in seconds in each type of blocks) and overall efficiency score (calculated by dividing percentage of accurate responses by mean RTs in seconds in all blocks).

A 2 (block: pure efficiency score versus mixed efficiency score) x 2 (language group: monolinguals versus bilinguals) mixed ANOVA with block as a within subjects factor and language group as a between subjects factor revealed a significant effect of block ($F(1, 59)=28.581$, $p<0.0001$), indicating better performance in the pure blocks and a significant effect of group indicating better multilingual performance overall ($F(1, 59)=4.013$, $p<0.050$). The critical block x group interaction was not significant ($F(1, 59)=0.040$, $p>0.05$), indicating that the bilingual advantage was in overall performance in the task and not due to better performance in the mixed blocks where the ability to switch between tasks is required. Further ANOVAs revealed no significant differences between the two groups for the mixing cost and switching cost measures (although there was a trend for a bilectal advantage for the switching cost measure; ($F(1, 59)=3.016$, $p>0.05$) and ($F(1, 59)=0.547$, $p>0.05$) respectively).

3.3.4. Backward Digit Span Task.

Mean scores for each language group are reported in table 3. An ANOVA revealed no significant differences between the two groups ($F(1, 59)=2.260$, $p>0.05$), although there was a numerical trend towards a bilingual advantage.

3.3.5. Corsi Blocks Task.

Mean scores per language group and condition (forward versus backward) are reported in table 3. A 2 (condition: forward versus backward) x 2 (language group: bilectals versus multilinguals) mixed ANOVA with condition as a within subjects factor and language group as a between subjects factor revealed only a significant effect of condition ($F(1, 62)=39.143$, $p<0.0001$) indicating better performance in the forward condition.

Table 3. Means and standard deviations per dependent measure (in each task) and language group.

Task	Measure	Bilectals		Multilinguals	
		Mean	SD	Mean	SD
Simon Task	RT Incongruent	1048.70	300.18	878.16	159.31
	RT Congruent	950.37	265.22	828.49	175.39
	RT Center	861.64	180.29	783.68	146.64
	Simon effect	98.34	98.46	50.67	82.74
Soccer Task	RT Go	1219.48	390.28	1224.45	310.86
	Delay	438.16	216.02	474.66	182.50
	SSRT	781.71	252.06	749.79	179.40
	Accuracy (%)	58.88	26.89	69.79	15.68
Color-Shape Task	RTs Repeat trials (Mixed blocks)	2835	1471.42	2335.77	729.78
	RTs Switch trials (Mixed Blocks)	2650.89	1137.62	2378.66	722.15
	RTs Pure Blocks	2015.30	991.38	1743.50	505.07
	RTs Mixed Blocks	2284.80	935.28	1845.20	505.44
	Accuracy (%) (Pure Blocks)	87.04	12,38	90.97	11.26
	Accuracy (%) (Mixed Blocks)	84.26	10.15	85.11	12.19
	Switching Cost	-184.42	595.22	49.89	368.01
	Mixing Cost	730.06	723.52	592.24	632.40
	Efficciency Score (Pure Blocks)	47.89	19.04	55.67	14.36
	Efficiency Score (Mixed Blocks)	41.43	14.49	48.71	11.87
	Efficiency Score (Total)	44.66	16.20	52.19	12.26
BDST	Total Score	4.13	1.13	4.63	1.54
Corsi Blocks Task	Forward Total Score	6.29	1.83	7.03	1.56
	Backward Total Score	5.029	1.75	5.47	2.18
Frith-Happe Animations Task	Random	3.46	0.70	3.55	0.78
	GD	2.60	1.17	2.76	1.06
	TOM	2.91	0.95	2.79	1.18
	MCQ-feelings	3.46	0.70	3.55	0.78

3.4.Theory of Mind Task.

Mean scores per language group and condition (ToM, GD, Random, MCQ-feelings) are presented in table 3. Mauchly's test indicated that the assumption of sphericity had been violated, W=.987, $p<.0001$, therefore, for the repeated measures effects and their interactions, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity (ε=.99). A 3 (condition: ToM, GD, Random) x 2 (language group: monolinguals versus bilinguals) mixed ANOVA with condition as a within subjects factor and language group as a between subjects factor revealed only a significant effect of condition ($F(2, 124)=16.435$, $p<0.0001$). Post-hoc pairwise comparisons with Bonferroni correction for multiple comparisons revealed the following results: performance in the *random* condition (mean=3.5) of the task was significantly different from performance in both the GD (mean=2.68) and ToM (mean=2.85) conditions (both *p*s<.05). The difference between performance in the GD and ToM conditions was not significant ($p>0.05$). ANOVAs conducted with MCQ-feelings or the combined MCQ-feelings and TOM score as dependent measures and language group as a between subjects factor revealed no significant differences between the two groups ($F(1, 61)=.690$, $p>0.05$ and ($F(1, 58)=1.286$, $p<0.05$) respectively).

3.5. The relation between conversational ability and the cognitive tests administered.

Of particular interest were the relations between the various cognitive factors tested and performance in the conversational test. We ran correlation analyses with overall performance in the conversational test and each type of implicatures as dependent measures and the various cognitive and background measures tested as independent variables. Overall performance in the conversational test was calculated as the sum of correct responses in all sub-tests (6 sub-tests x 3 items each resulting in a maximum total score of 18). We also used a combined score for the two sub-tests on scalar implicatures since both sub-tests tested comprehension of the same type of implicature. This score was calculated as the sum of correct responses in the two sub-tests (Scalars total). Correlation analyses were conducted on the full set of data including the 9 participants that were excluded from the previous analyses. The following dependent measures from each task were included in the correlation analyses as independent variables: Simon effect (Simon task), SSRT (soccer task), mixing cost (Color-Shape task), total scores in the BDST, forward (Corsi forward) and backward version (Corsi backward) of the Corsi blocks task, the ToM-MCQ-feelings score (Frith-Happe animations task) expressive vocabulary total score, IQ score, gender, age in years and version of conversational test (version).

For all correlation analyses we followed the following procedure: first we looked at the two-tailed bivariate correlations between each dependent measure and the independent variables. Then multiple regression analyses were conducted using the forced entry method in SPSS. Multiple regression analyses included all independent variables that significantly correlated with the dependent measure and also any independent variables that significantly correlated with these in order to partial out the shared variance between these variables.

These analyses revealed the following results: The overall conversational score significantly and positively correlated with Corsi forward (r(two-tailed)=.255, $p<.05$) and age (r(two-tailed)=.275, $p<.05$) and negatively correlated with version (r(two-tailed)=-.356, $p<.05$). Corsi forward significantly correlated with Corsi backward (r(two-tailed)=.461, $p<.05$), IQ (r(two-tailed)=.393, $p<.05$), BDST (r(two-tailed)=.323, $p<.05$) and ToM-MCQ-feelings (r(two-tailed)=.294, $p<.05$). Age significantly correlated with IQ (r(two-tailed)=.471, $p<.05$) BDST (r(two-tailed)=.305, $p<.05$), Corsi Forward (r(two-tailed)=.497, $p<.05$), Corsi Backward (r(two-tailed)=.387, $p<.05$) and ToM-MCQ-feelings (r(two-tailed)=.371, $p<.05$). IQ significantly correlated with age (r(two-tailed)=.471, $p<.05$), Corsi forward (r(two-tailed)=.497, $p<.05$), Corsi Backward (r(two-tailed)=.387, $p<.05$) and ToM-MCQ-feelings (r(two-tailed)=.371, $p<.05$). A multiple regression with Corsi forward, age, version, Corsi backward, IQ, BDST and ToM-MCQ-feelings as independent variables revealed a significant effect of age (t(59)=2.400, p<.05) and a significant effect of version (t(59)=13.859, p<.05). The overall model was significant (F(7.66)=3.306, p<.05) accounting for 28.2% of total variance in the dependent measure.

The Scalars total score significantly correlated with Corsi forward (r(two-tailed)=.297, p<.05), age (r(two-tailed)=.369, p<.05) and IQ (r(two-tailed)=.240, p<.05). Corsi forward significantly correlated with Corsi backward (r(two-tailed)=.461, p<.05), IQ (r(two-tailed)=.393, p<.05), BDST (r(two-tailed)=.323, p<.05) and ToM-MCQ-feelings (r(two-tailed)=.294, p<.05). IQ significantly correlated with age (r(two-tailed)=.471, p<.05), Corsi forward (r(two-tailed)=.497, p<.05), Corsi Backward (r(two-tailed)=.387, p<.05) and ToM-MCQ-feelings (r(two-tailed)=.371, p<.05). A multiple regression with Corsi forward, age, IQ, Corsi backward, BDST and ToM-MCQ-feelings as independent variables revealed only a marginally significant effect of age (t(59)=1.934, p=.058). The overall model was marginally significant (F(7.66)=1.970, p=.084).

Metaphor significantly correlated with Corsi backward (r(two-tailed)=.239, p<.05), age (r(two-tailed)=.273, p<.05) and version (r(two-tailed)=-.260, p<.05). Corsi backward significantly correlated with Corsi forward (r(two-tailed)=.461, p<.05), IQ (r(two-tailed)=.346, p<.05) and BDST (r(two-tailed)=.302, p<.05). Age significantly correlated with IQ (r(two-tailed)=.471, p<.05) BDST (r(two-tailed)=.305, p<.05), Corsi Forward (r(two-tailed)=.497, p<.05), Corsi Backward (r(two-tailed)=.387, p<.05) and ToM-MCQ-feelings (r(two-tailed)=.371, p<.05). A multiple regression analysis with Corsi backward, age, version, Corsi forward, IQ, BDST and ToM-MCQ-feelings as independent variables revealed a marginally significant effect of age (t(59)=1.880, p=.065) and a significant effect of version (t(59)=-3.404, p<.05). The overall model was significant (F(7.66)=3.306, p<.05) accounting for 28.2% of total variance in the dependent measure.

Irony significantly correlated only with version (r(two-tailed)=-.510, p<.05) and manner implicatures did not correlate with any independent variable. Therefore, no regression analyses were conducted for these dependent measures.

4. Discussion.

The aims of the present investigation were to examine (1) the effect of multilingual experience on children's comprehension of implicatures and (2) the relations between pragmatic competence and cognitive factors such as EC and ToM.

The results of our experiment provide only limited support for the position that multilingualism fosters children's pragmatic abilities. Overall, multilingual children were numerically better than bilectals in most implicature sub-tests including metaphor, irony, scalars 1 and scalars 2. Bilectals were numerically better in comprehending manner (but see the complex interaction pattern in section 3.2.) and relevance implicatures, however the differences between the two groups for these two sub-tests were minimal while performance in the relevance sub-test was almost at ceiling for both groups. The only statistically significant result indicated a multilingual advantage in the comprehension of irony but only in version 1 of the conversational test. While the overall pattern of our results points towards a multilingual advantage, especially for the more demanding types of implicatures (i.e. irony), it does not provide robust evidence in favor of the hypothesis that exposure to more than one languages confers an advantage in children's conversational understanding.

Having said that, three previous studies conducted by Siegal and colleagues (2007; 2009; 2010) provided strong evidence for a bilingual advantage in pragmatic ability. What could be driving this variability in the findings? To start with, as already noted in previous sections it is possible that the bilingual advantage reported in Siegal et al. (2009; 2010) did not indicate advanced pragmatic ability, as the CVT employed in their studies included items that confounded pragmatic ability with world knowledge. Nevertheless, Siegal and colleagues did report a bilingual advantage for the items on the maxims of manner (Siegal et al. 2009; 2010) and quantity (Siegal et al. 2007; 2010) which unequivocally test pragmatic sensitivity. Moreover, our study departed from those of Siegal et al. (2007; 2009; 2010) in several respects. First, Siegal et al. (2009; 2010) tested children's sensitivity to pragmatic violations while we tested children's comprehension of implicatures. It could be the case that a bilingual advantage in detecting pragmatic violations does not extend to the ability to understand more complex pragmatic language like implicatures or, alternatively, that the effect of bilingualism on implicature comprehension is simply smaller and thus more difficult to detect. Nevertheless, the results of the Siegal et al. (2007) study seem to refute this explanation, as they showed a bilingual advantage in children's ability to comprehend scalar implicatures using a task that closely resembled one of the tasks used in our study (scalars 2). Second, Siegal et al. (2007; 2009; 2010) tested children of a younger age (preschoolers between 4–6 year olds) than the children recruited for the current study (school-aged children between 6–12 years old). It is possible that the impact of bilingualism is more profound during the preschool years, a period during which children make impressive strides in their social-pragmatic cognitive development. Again, however, we see no reason for why a bilingual effect might not show up for older children: pragmatic development does not stop after the preschool years and indeed more complex forms of pragmatic understanding (e.g. irony, metaphor) continue to develop until late childhood. That said, some types of implicatures tested in our study were very easy to understand and thus revealed no or limited variability in children's performance: relevance implicatures were invariably understood by all children whereas comprehension of scalar implicatures in sub-test 1 showed variability mostly in the younger age group tested (6- to 7-year-olds). It is, thus, likely that potential differences between the two language groups do exist in these types of implicatures and that only a more sensitive study using younger preschool-aged children will reveal.

Finally, one could argue that the lack of robust evidence in favor of a multilingual advantage in pragmatics in our sample was due to the special language profile of children growing up in the Republic of Cyprus. The control group in our study consisted of bilectal children being exposed to SMG mostly through schooling and the media and to CG mainly through their everyday life. Multilinguals on the other hand were children with extensive exposure to CG and another language through schooling and everyday social interactions and with limited exposure to SMG through a few hours of formal teaching at school (and possibly through the media). Although bilingualism has been shown to have a positive effect on pragmatic abilities it is not at all clear what the effect of bilectalism is on these abilities. Siegal et al. (2009) suggest that the source of the bilingual advantage in pragmatics might be a compensatory mechanism that enhances bilinguals' pragmatic skills as an offset to their well-known delays in vocabulary and grammatical knowledge. Because of their weaker language skills bilingual children may become more attentive than monolinguals to contextual aspects of the communicative situation and use this ability in order to infer the speakers' meanings. If CG-speaking bilectal children face similar difficulties in their SMG language abilities then it is possible that the same compensatory mechanism works to enhance their pragmatic skills although perhaps to a smaller degree. If indeed this is the case for bilectals then one would expect either a smaller effect of multilingualism or no effect at all. Of course, in the absence of a true monolingual control group in our study it is not possible to say whether there is no multilingual effect at all or whether we failed to replicate it in our study due to the bilectal status of our control group. Future research may shed more light on this issue.

Our results also revealed that the regular use of more than one languages leads to an advancement in children's inhibitory control skills. Although faster in all conditions of the

Simon task (surprisingly including the neutral condition where no EC abilities are involved) multilingual children exhibited a significantly smaller Simon effect than bilectal children indicating better ability to suppress task-irrelevant information during incongruent trials. Furthermore, our study showed that the multilingual advantage in EC appears to be isolated to executive control tasks that require inhibition, specifically the ability to inhibit task-irrelevant conflicting information, but not cognitive flexibility or working memory. Moreover, we failed to replicate the finding of a bilingual advantage in children's ToM abilities. Could this failure to find a multilingual effect on ToM or on some of the EC measures be due to the bilectal language profile of our control group? We find this highly unlikely since we found a multilingual advantage in children's inhibitory control abilities. This suggests that, whatever (if any) the effect of bilectalism is on children's cognitive profile, speaking two or more languages has an additional effect. Thus if multilingualism indeed had an impact on these cognitive skills, this effect ought to have materialized in our study in the same way as the effect on children's inhibitory control skills did.

Finally, we found no evidence for a positive link between pragmatic ability and cognitive factors such as EC or ToM. The same null result was also obtained in all three studies by Siegal and colleagues (2007; 2009; 2011). However, their studies were limited in that they tested sensitivity to pragmatic violations, they employed only two measures of EC, they did not use any measures of working memory or ToM and their EC tasks showed a ceiling or near ceiling effect of performance. In this study, we used a wider range of cognitive measures corresponding to various aspects of EC and ToM. We also tested the ability to comprehend implicatures (instead of sensitivity to pragmatic violations) for which there are both theoretical and empirical grounds to suggest that is more dependent on cognitive resources. All of our measures showed adequate variability and one of our measures was sensitive enough to reveal significant differences between the two language groups in our study. In this respect the evidence from our experiment provide no support for the view that EC or ToM underpin the development of pragmatic ability.

Overall, while the widely reported bilingual advantage in inhibitory control was indeed replicated in our sample, there was only suggestive evidence for an advantage in pragmatic language when using a task that (a) unambiguously tests for pragmatic ability (rather than world knowledge), and (b) studies implicature comprehension rather than mere detection of pragmatic violations. Moreover, no link emerged between executive functions and pragmatic competence. Additional research is also required in order to establish the consequences of bilectalism on children's language, pragmatic and cognitive performance. Indeed the failure to find robust evidence of a multilingual advantage in pragmatics in our study might be related to the bilectal status of our control group.

References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207–245.

Antoniou, K., & Katsos, N. (2011). Only some adults reject under-informative utterances. Poster presented at the 4th Biennial Conference of Experimental Pragmatics. Barcelona.

Arvaniti, A. (2010). Linguistic practices in Cyprus and the emergence of Cypriot Standard Greek. *Language Review, 17*, 15–45.

Bernicot, J., Laval, V., & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics, 39*(12), 2115–2132.

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

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*: Cambridge University Press.

Bialystok, E. (2009). Claiming evidence from non-evidence: a reply to Morton and Harper. *Developmental science, 12*(4), 499–501.

Bialystok, E. (2010). Global–local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Developmental psychology, 46*(1), 93.

Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. *Journal of experimental child psychology, 110*(3), 461–468.

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., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest, 10*(3), 89–129.

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

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition, 112*(3), 494.

Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2010). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental science, 14*(2), 256–269.

Borke, H. (1971). Interpersonal perception of young children: Egocentrism or empathy? *Developmental psychology, 5*(2), 263.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language, 51*(3), 437–457.

Bretherton, I., Fritz, J., Zahn-Waxler, C., & Ridgeway, D. (1986). Learning to talk about emotions: A functionalist perspective. *Child development*, 529–548.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic bulletin & review, 16*(5), 893–900.

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

Corsi, P. M. (1973). *Human memory and the medial temporal region of the brain.* Unpublished doctoral dissertation, McGill University.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*(2), 135–149.

Crystal, D. (2003). *English as a Global Language*: Cambridge University Press.
de Abreu, P. M. E. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19*(5), 529–537.
de Abreu, P. M. E., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism Enriches the Poor Enhanced Cognitive Control in Low-Income Minority Children. *Psychological science, 23*(11), 1364–1371.
De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie), 54*(2), 128–133.
Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *The Quarterly Journal of Experimental Psychology, 64*(12), 2352–2367.
Ellefson, M. R., Shapiro, L. R., & Chater, N. (2006). Asymmetrical switch costs in children. *Cognitive Development, 21*(2), 108–130.
Foy, J. G., & Mann, V. A. (2013). Bilingual children show advantages in nonverbal auditory executive function task. *International Journal of Bilingualism*, doi: 10.1177/1367006912472263.
Gathercole, V.C.M. (2002a). Command of the mass/count distinction in bilingual and monolingual children. An English morphosyntactic distinction. In *Language and Literacy in Bilingual Children,* ed. D.K. Oller and R.E. Eilers, 175–206. Clevedon, UK: Multilingual Matters.
Gathercole, V. C. M. (2002b). Grammatical gender in bilingual and monolingual children: A Spanish morphosyntactic distinction. In *Language and Literacy in Bilingual Children,* ed. D.K. Oller and R.E. Eilers, 207–219. Clevedon, UK: Multilingual Matters.
Gathercole, V. C. M. (2002c). Monolingual and bilingual acquisition: Learning different treatments of that-trace phenomena in English and Spanish. In *Language and Literacy in Bilingual Children,* ed. D.K. Oller and R.E. Eilers, 175–206. Clevedon, UK: Multilingual Matters.
Genesee, F., Tucker, G. R., & Lambert, W. E. (1975). Communication skills of bilingual children. *Child development*, 1010–1014.
Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism Language and Cognition., 6*(1), 1–16.
Greenberg, A., Bellana, B., & Bialystok, E. (2012). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive Development*, 28, 41–50.
Grice, H. P. (1989). *Studies in the Way of Words*: Harvard University Press.
Grosjean, F. (2010). *Bilingual: Life and Reality*: Harvard University Press.
Hilchey, M. D., & Klein, R. M. (2011a). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic bulletin & review, 18*(4), 625–658.
Hill, E. L. (2004). Executive dysfunction in autism. *Trends in cognitive sciences, 8*(1), 26–32.
Junker, D. A., & Stockman, I. J. (2002). Expressive vocabulary of German-English bilingual toddlers. *American Journal of Speech-Language Pathology, 11*(4), 381.
Katsos, N., Roqueta, C. A., Estevan, R. A. C., & Cummins, C. (2011). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification? *Cognition, 119*(1), 43–57.
Keil, F. C. (1986). Conceptual domains and the acquisition of metaphor. *Cognitive Development, 1*(1), 73–96.
Kovács, Á. M. (2008). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental science, 12*(1), 48–54.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Mit Press.

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. *Inhibitory processes in attention, memory and language*, 189–239.

Loukusa, S., & Moilanen, I. (2009). Pragmatic inference abilities in individuals with Asperger syndrome or high-functioning autism. A review. *Research in Autism Spectrum Disorders, 3*(4), 890–904.

Marinis, T., & Chondrogianni, V. (2010). Production of tense marking in successive bilingual children: When do they converge with their monolingual peers? *International journal of speech-language pathology, 12*(1), 19–28.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism, Language and Cognition, 11*(1), 81.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., & Wagner, T.D. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: a latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Morales, J., Calvo, A., & Bialystok, E. (2012). Working memory development in monolingual and bilingual children. *Journal of experimental child psychology, 114,* 187–202

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

Moses, L. J., Carlson, S. M., & Sabbagh, M. A. (2005). On the specificity of the relation between executive function and children's theories of mind. In *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind*, ed. W. Schneider, R. Schumann-Hengsteler & B. Sodian, 131–145. Clevedon, UK: Multilingual Matters.

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58*(2), 220–249.

Oller, D. K., & Eilers, R. E. (2002). *Language Literacy in Bilingual Children*. Clevedon, UK: Multilingual Matters.

Ozonoff, S. (1997). Components of executive function in autism and other disorders. In *Autism as an executive disorder,* ed. J. Russel, 179–211.

Oxford: Oxford University Press. Pearson, B. Z., Fernandez, S., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of child language, 22*, 345–345.

Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (2006). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language learning, 43*(1), 93–120.

Bialystok, E., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525–531.

Petitto, L.-A., & Kovelman, I. (2003). The bilingual paradox: How signing-speaking bilingual children help us to resolve it and teach us about the brain's mechanisms underlying all language acquisition. *Learning Languages, 8*(3), 5–18.

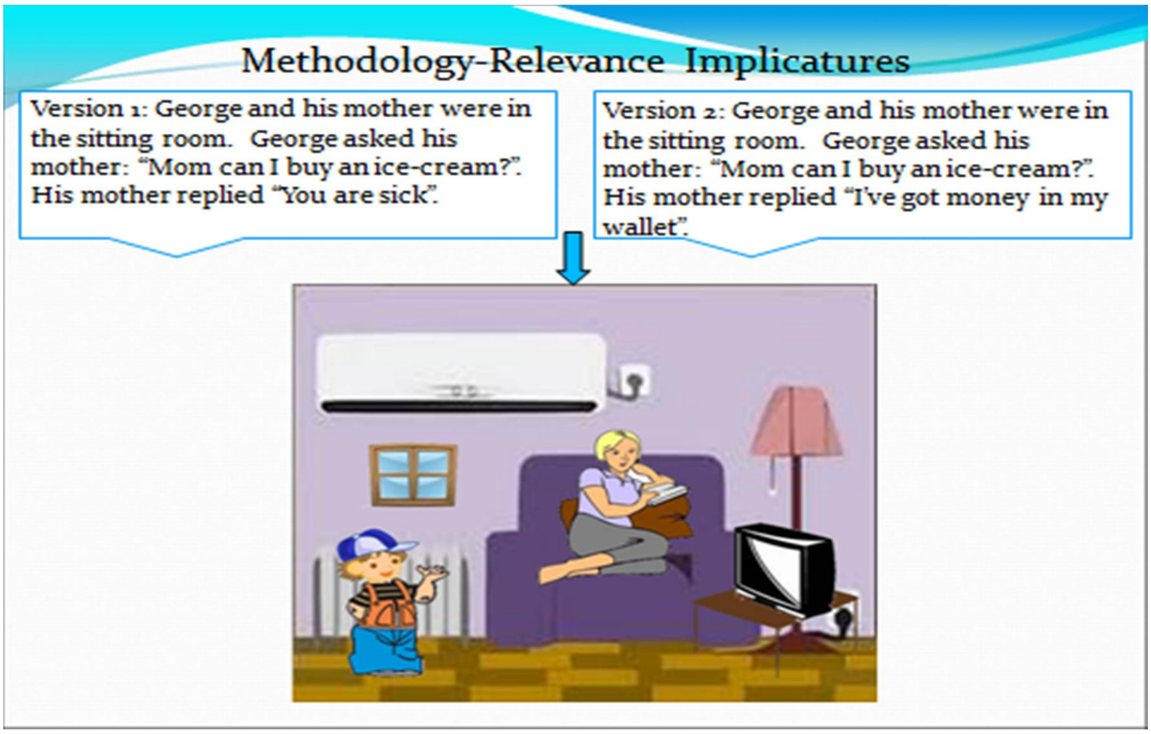
Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of experimental child psychology, 113,* 535–551.

Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2012). Lexical access and vocabulary development in very young bilinguals. *International Journal of*

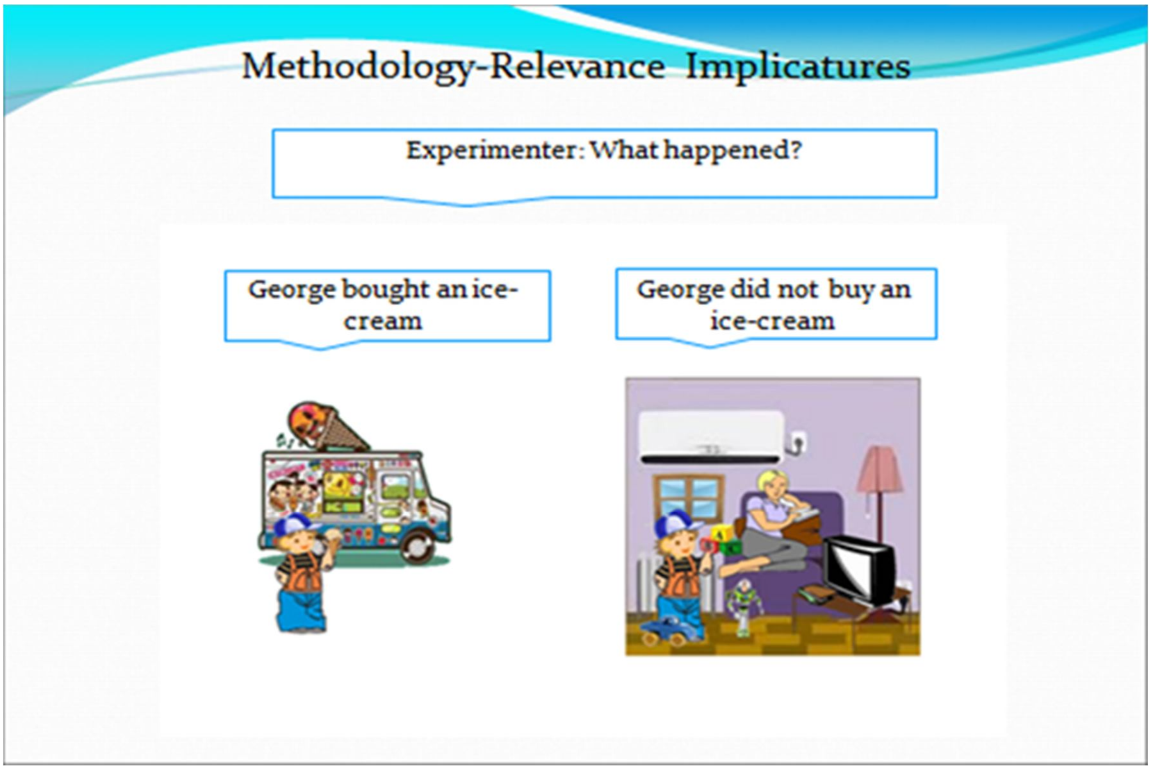
Bilingualism, *17*(1), 57–70.
Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of experimental child psychology, 108*(3), 567–579.
Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language acquisition, 14*(4), 347–375.
Rowe, C. & Grohmann, K.K. (under review). Discrete bilectalism: Towards co-overt prestige and diglossic shift in Cyprus.
Schwartz, R. M., Trabasso, T., Izard, C., Kagan, J., & Zajonc, R. (1984). Children's understanding of emotions. *Emotions, cognition, and behavior*, 409–437.
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,* ed. Y. Otsu. Tokyo: Hituzi Syobo.
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.
Simon, J. R. (1969). Reactions towards the source of stimulation. *Journal of experimental psychology*, *81*, 174–176.
Smiley, P., Huttenlocher, J., Saarni, C., & Harris, P. (1984). Young children's understanding of facial and situational components of emotion words. *Unpublished manuscript, University of Chicago*.
Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind & Language, 17*(1-2), 3–23.
Stein, N. L., & Levine, L. J. (1987). Thinking about feelings: The development and organization of emotional knowledge. *Aptitude, learning, and instruction, 3*, 165–197.
Vogindroukas, I., Protopapas, A. & Sideridis, G. (2009). Test of Expressive Vocabulary [in Greek]. Chania: Glauki.
Vosniadou, S. (1987). Children and metaphors. *Child development*, 870–885.
Waggoner, J. E., & Palermo, D. S. (1989). Betty is a bouncing bubble: Children's comprehension of emotion-descriptive metaphors. *Developmental psychology, 25*(1), 152.
Waggoner, J. E., Palermo, D. S., & Kirsh, S. J. (1997). Bouncing bubbles can pop: Contextual sensitivity in children's metaphor comprehension. *Metaphor and Symbol, 12*(4), 217–229.
Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. San Antonio, TX, US: Psychological Corporation.
Winner, E. (1988). The point of words: *Children's understanding of metaphor and irony*. Harvard University Press.
Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition, 14*(3), 412–422.
Yow, W. Q., & Markman, E. M. (2011). Young bilingual children's heightened sensitivity to referential cues. *Journal of Cognition and Development, 12*(1), 12–31.

Appendix

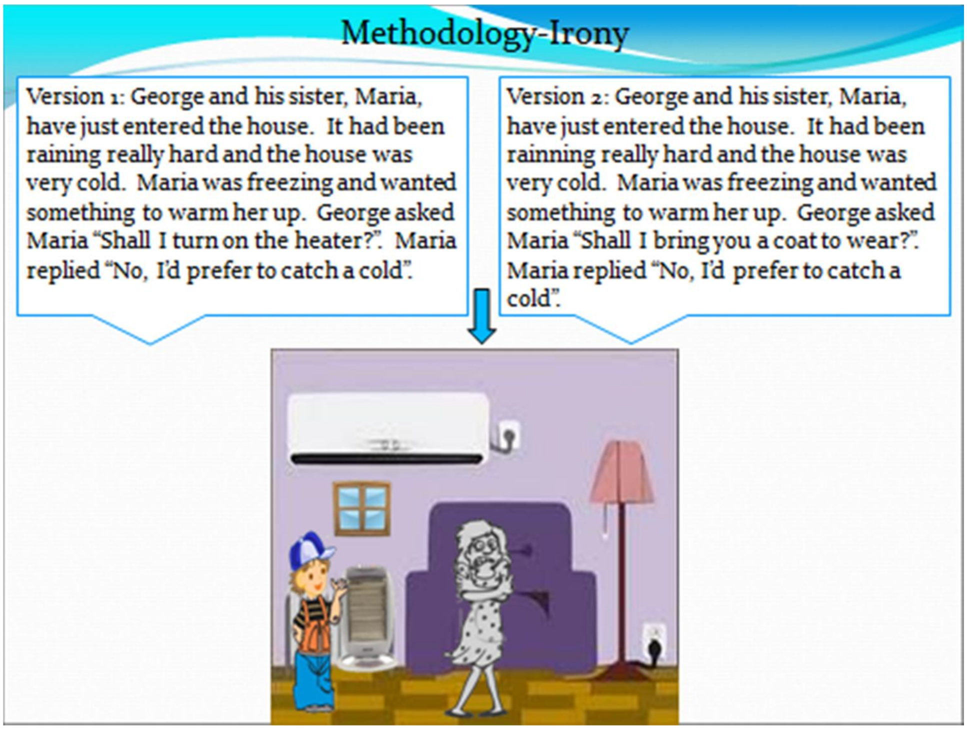
(1)


Methodology-Relevance Implicatures
Version 1: George and his mother were in the sitting room. George asked his mother: "Mom can I buy an ice-cream?". His mother replied "You are sick".
Version 2: George and his mother were in the sitting room. George asked his mother: "Mom can I buy an ice-cream?". His mother replied "I've got money in my wallet".

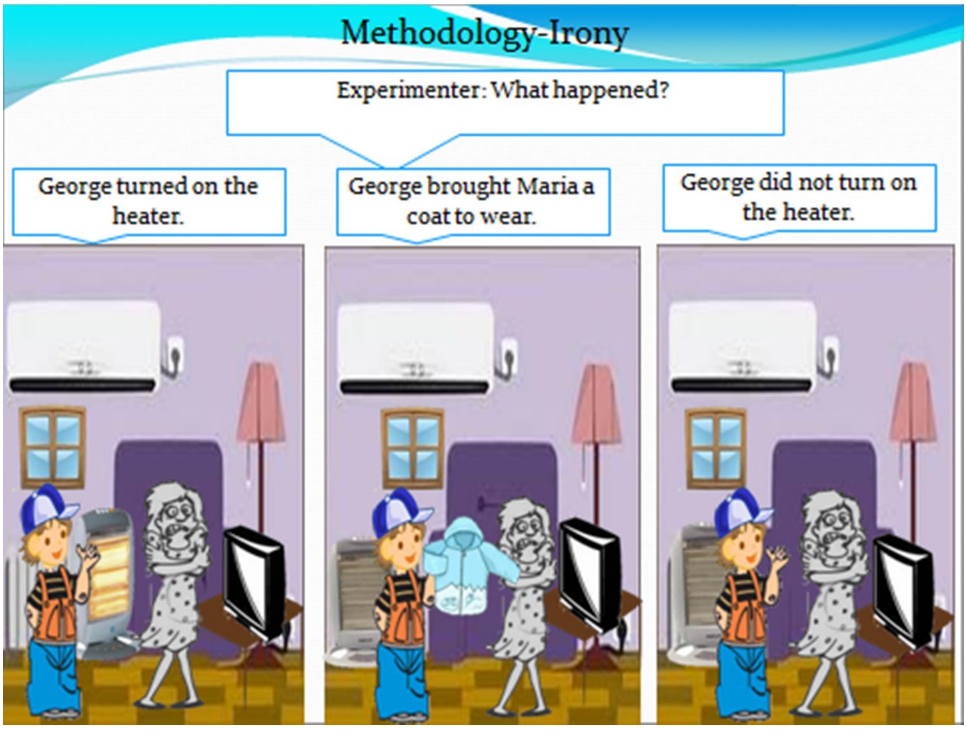
(2)


Methodology-Relevance Implicatures
Experimenter: What happened?
George bought an ice-cream
George did not buy an ice-cream

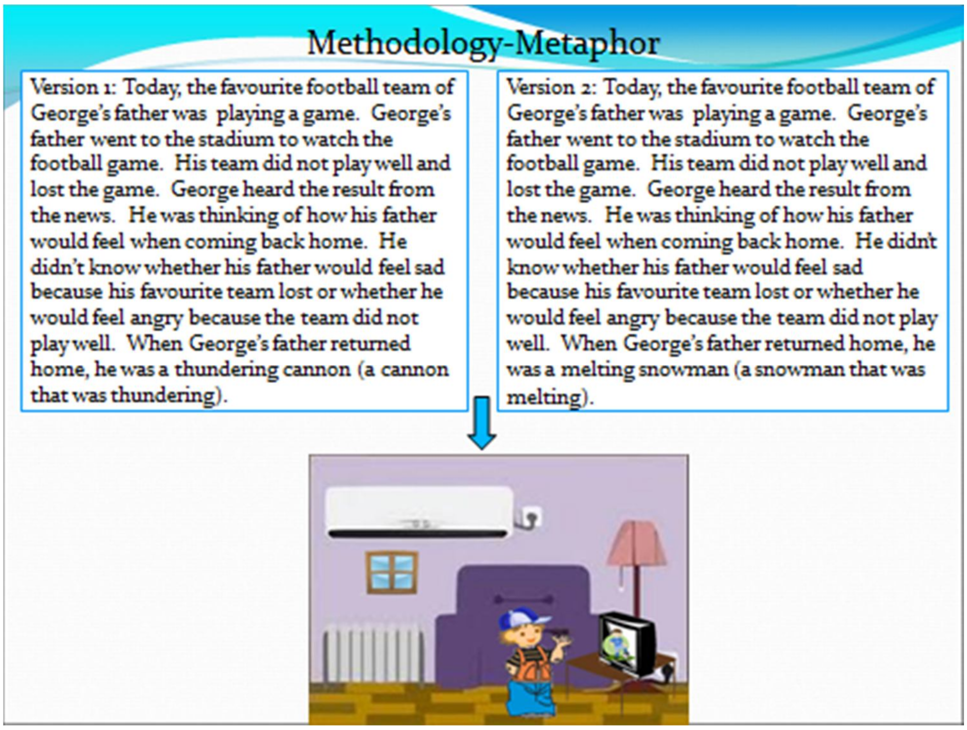
(3)


Methodology-Irony
Version 1: George and his sister, Maria, have just entered the house. It had been raining really hard and the house was very cold. Maria was freezing and wanted something to warm her up. George asked Maria "Shall I turn on the heater?". Maria replied "No, I'd prefer to catch a cold".
Version 2: George and his sister, Maria, have just entered the house. It had been rainning really hard and the house was very cold. Maria was freezing and wanted something to warm her up. George asked Maria "Shall I bring you a coat to wear?". Maria replied "No, I'd prefer to catch a cold".

(4)


Methodology-Irony
Experimenter: What happened?
George turned on the heater.
George brought Maria a coat to wear.
George did not turn on the heater.

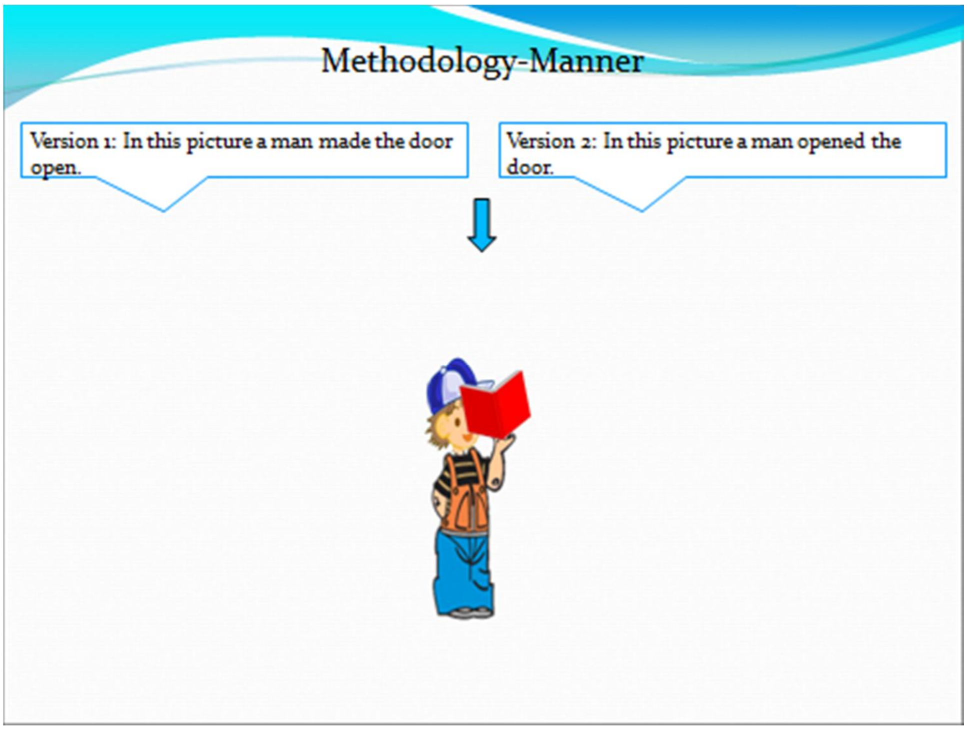
(5)


Methodology-Metaphor
Version 1: Today, the favourite football team of George's father was playing a game. George's father went to the stadium to watch the football game. His team did not play well and lost the game. George heard the result from the news. He was thinking of how his father would feel when coming back home. He didn't know whether his father would feel sad because his favourite team lost or whether he would feel angry because the team did not play well. When George's father returned home, he was a thundering cannon (a cannon that was thundering).
Version 2: Today, the favourite football team of George's father was playing a game. George's father went to the stadium to watch the football game. His team did not play well and lost the game. George heard the result from the news. He was thinking of how his father would feel when coming back home. He didn't know whether his father would feel sad because his favourite team lost or whether he would feel angry because the team did not play well. When George's father returned home, he was a melting snowman (a snowman that was melting).

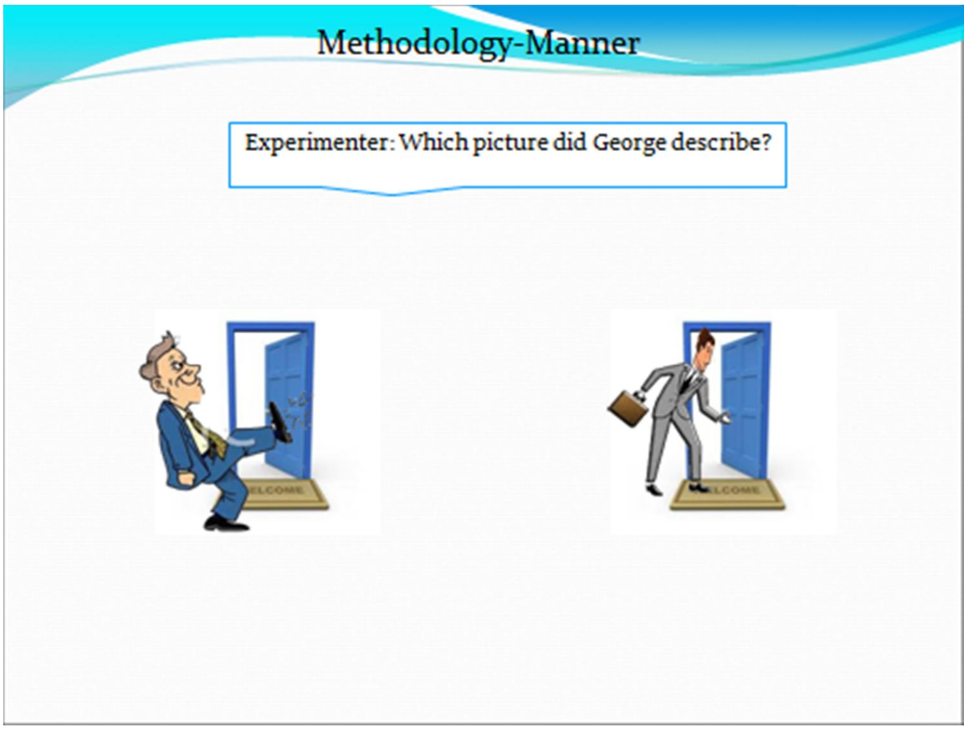
(6)


Methodology-Metaphor
Experimenter: How was George's father feeling?
He was feeling angry.
He was feeling sad.

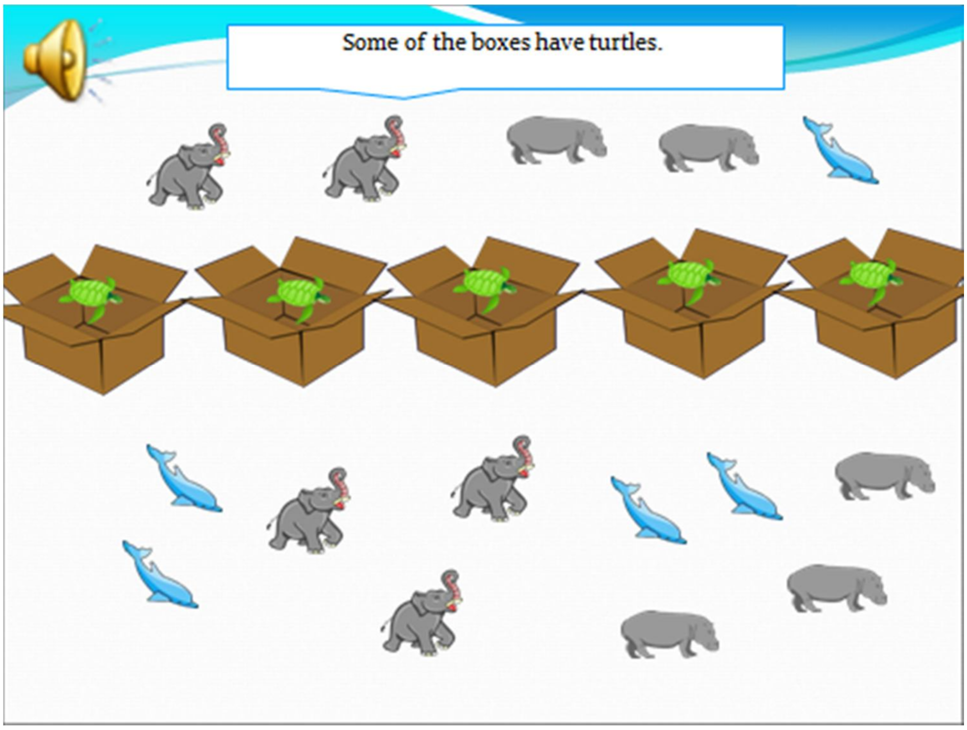
(7)


Methodology-Manner
Version 1: In this picture a man made the door open.
Version 2: In this picture a man opened the door.

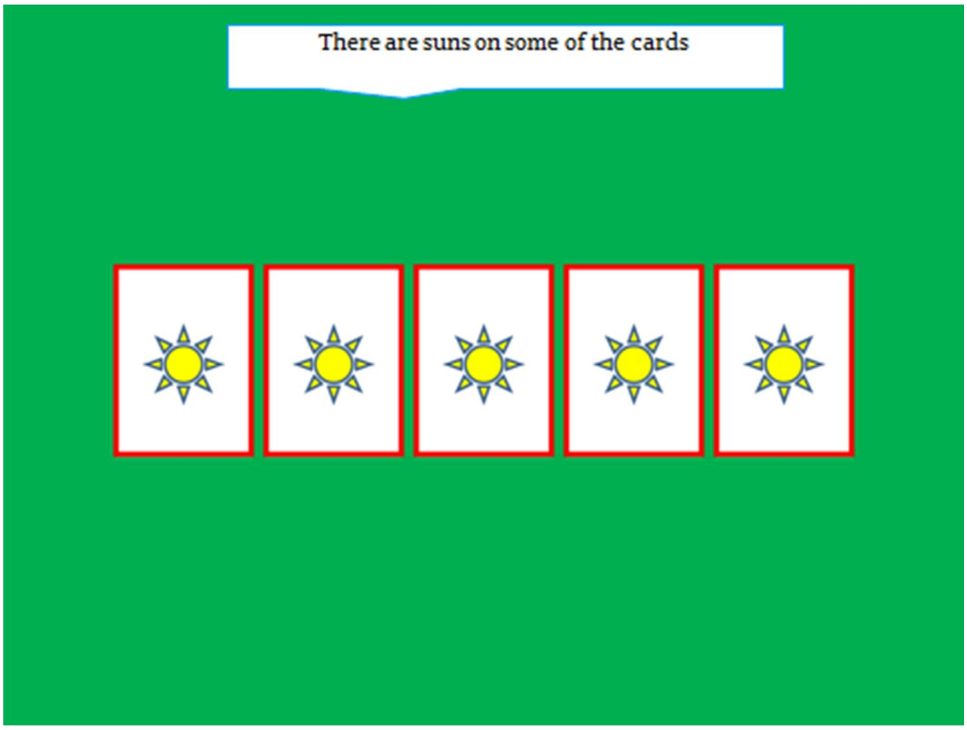
(8)


Methodology-Manner
Experimenter: Which picture did George describe?

(9)


Some of the boxes have turtles.

(10)


There are suns on some of the cards